Quantum Mathematics

by Peter J. Olver
University of Minnesota

Table of Contents

I think I can safely say that nobody understands quantum mechanics.

<div align="right">— Richard P. Feynman, [22].</div>

1. Introduction.

These are the lecture notes that I prepared for a graduate level course on the basics of mathematical quantum mechanics that I gave in the academic year 1989–90 at the University of Minnesota. The original notes were typed in Microsoft Word and hence never distributed widely. After 15 years elapsed, I was finally inspired to convert them to TEX. Aside from minor editing, corrections, reordering of some of the material, and updating the references, they are the same as were distributed to the students in the class.

The goal was to provide a broad overview of mathematical techniques used in quantum mechanics. While applications appear throughout, the emphasis is on the fundamental mathematical tools used in the physical theory. Techniques include Hamiltonian mechanics, quantization, operator theory, stationary phase, inverse scattering and solitons, group representation theory, Clifford algebras, and much more.

I hope that you, as reader, will find them of use and as enjoyable as when I first prepared them and gave the course.

2. Hamiltonian Mechanics.

<div align="right">References: [2], [5], [17], [25], [27], [42], [58].</div>

The fundamental insight of Schrödinger's approach to quantum mechanics was that the quantum mechanical counterparts to the equations of classical mechanics are most easily found by writing the latter in Hamiltonian form. Indeed, Hamilton himself could, in direct analogy with the correspondence between geometric optics and wave optics — the prototype of wave/particle duality — have straightforwardly written down the equations of "wave mechanics", and thus anticipated quantum mechanics by almost a century. However, the lack of any physical motivation for taking this conceptual leap prevented such a mathematical advance occurring before its time. We thus begin our study with an overview of basic Hamiltonian mechanics.

Hamiltonian Systems

We consider a system of first order ordinary differential equations on the *phase space* $M = \mathbb{R}^{2n}$, with *canonical local coordinates* $(p, q) = (p_1, \ldots, p_n, q_1, \ldots, q_n)$. The system is called *Hamiltonian* if there is a smooth function $H(p, q)$ such that it can be written in the form

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad i = 1, \ldots, n. \qquad (2.1)$$

Typically the q's represent the coordinates of the particles in the system, the p's the corresponding momenta, and H is the physical energy. Essentially, Hamilton's equations provide a convenient canonical form for writing the equations of conservative classical mechanics.

A good example is the motion of masses in a potential force field. Here the q's represent the positions of the masses. Let $U(t, q)$ be the potential function, so the corresponding force

field is its negative spatial gradient: $F = -\nabla U = (-\partial U/\partial q_1, \ldots, -\partial U/\partial q_n)$. Newton's equations are

$$m_i \frac{d^2 q_i}{dt^2} = -\frac{\partial U}{\partial q_i}, \qquad i = 1, \ldots, n,$$

where m_i is the ith mass. (For three-dimensional motion, $n = 3m$, and the positions and masses are labelled accordingly.) To put these equations in Hamiltonian form, we let

$$p_i = m_i \dot{q}_i,$$

be the ith momentum, and the Newtonian dot notation is used for time derivatives: $\dot{q} = dq/dt$. We define the Hamiltonian function

$$H = \frac{1}{2} \sum_{i=1}^{n} m_i \dot{q}_i^2 + U(t, q) = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + U(t, q)$$

to be the total energy (kinetic + potential). Then Hamilton's equations (2.1) are

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i}, \qquad \frac{dq_i}{dt} = \frac{p_i}{m_i}, \qquad i = 1, \ldots, n,$$

and are clearly equivalent to Newton's equations of motion.

The system of differential equations governing the solutions to any (first order) variational problem in one independent variable can always be put into Hamiltonian form. Indeed, consider the problem of minimizing

$$\mathcal{I}[q] = \int_a^b L(t, q, \dot{q}), dt \qquad q(a) = a, \qquad q(b) = b, \tag{2.2}$$

where the *Lagrangian L* is a smooth function of t, q_i, and $\dot{q}_i = dq_i/dt$. Sufficiently smooth minima must satisfy the associated *Euler-Lagrange equations*

$$E_i(L) = -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} + \frac{\partial L}{\partial q_i} = 0, \qquad i = 1, \ldots, n. \tag{2.3}$$

To see this, suppose $q(t)$ is a smooth minimizer of (2.2). Then for smooth perturbations $h(t)$ satisfying $h(a) = h(b) = 0$ so as to preserve the boundary conditions, we must have

$$0 = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \mathcal{I}[q + \varepsilon h] = \int_a^b \sum_{i=1}^{n} \left[\frac{\partial L}{\partial q_i} h_i(t) + \frac{\partial L}{\partial \dot{q}_i}\frac{dh_i}{dt} \right] dt$$

$$= \int_a^b \sum_{i=1}^{n} \left[\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} \right] h_i(t) \, dt,$$

where we integrated the second set of summands by parts and used the boundary conditions on h to eliminate the boundary terms. Since the function $h(t)$ is arbitrary, an easy argument (the so-called duBois–Reymond Lemma) shows that the quantities in brackets must vanish, implying (2.3).

In general, we define the *momenta* via the *Legendre transformation*

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \tag{2.4}$$

and impose the nondegeneracy assumption

$$\det\left(\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}\right) \neq 0, \tag{2.5}$$

which, by the Implicit Function Theorem, allows us to locally solve (2.4) for $\dot{q}$ as a function of t, p, q:

$$\dot{q}_i = \varphi_i(t, p, q). \tag{2.6}$$

Define the *Hamiltonian function* to be

$$H(t, p, q) = \sum_{i=1}^{n} p_i \dot{q}_i - L(t, q, \dot{q}), \tag{2.7}$$

where we use (2.6) to replace $\dot{q}$. Then the Euler-Lagrange equations (2.3) are equivalent to Hamilton's equations (2.1). The easiest way to see this is to use differentials:

$$dH = \frac{\partial H}{\partial t} dt + \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i.$$

But, using (2.7), (2.4),

$$dH = p_i d\dot{q}_i + \dot{q}_i dp_i - \frac{\partial L}{\partial t} dt - \frac{\partial L}{\partial q_i} dq_i - \frac{\partial L}{\partial \dot{q}_i} d\dot{q}_i = \dot{q}_i dp_i - \frac{\partial L}{\partial t} dt - \frac{\partial L}{\partial q_i} dq_i.$$

Equating the latter two expressions, we deduce that

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}, \qquad \frac{\partial H}{\partial p_i} = \dot{q}_i, \qquad \frac{\partial H}{\partial q_i} = -\frac{\partial L}{\partial q_i} = -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} = -\dot{p}_i.$$

from which Hamilton's equations (2.1) follow immediately.

Example 2.1. In particle dynamics, the Lagrangian is the difference of kinetic and potential energy:

$$L = \frac{1}{2}\sum_{i=1}^{n} m_i \dot{q}_i^2 - U(t, q).$$

The Euler-Lagrange equations are just Newton's laws $F = ma$:

$$m_i \ddot{q}_i + \frac{\partial U}{\partial q_i} = 0, \qquad i = 1, \dots, n. \tag{2.8}$$

To place this system of second order ordinary differential equations in Hamiltonian form, in view of (2.4), set

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = m_i \dot{q}_i,$$

which is exactly the ith physical momentum. Thus, the Hamiltonian function (2.7) takes the form

$$H(p, q) = \frac{1}{2} \sum_{i=1}^{n} m_i \dot{q}_i^2 + U(t, q) = \sum_{i=1}^{n} \frac{p_i^2}{2 m_i} + U(t, q),$$

and is the total energy. As we have seen, Hamilton's equations (2.1) are equivalent to Newton's equations (2.8).

Example 2.2. *Geometric Optics*: According to *Fermat's Principle*, the light rays traveling between two points follows the path requiring the least time. If the light is traveling in a medium with index of refraction $n(q)$, which is the ratio of the speed of light in the medium to the speed of light in vacuo, c, then its velocity is $ds/dt = c/n$. Therefore, taking units in which $c = 1$, the time to transit between two points A and B is

$$t = \int_A^B dt = \int_A^B n(q)\, ds,$$

where ds denotes the element of arc length on the path travelled by the light. Fermat says that we need to minimize (or at least find stationary values for) this functional. Now ds is not a very convenient parametrization, as the endpoints of the functional are no longer fixed. However, the integral does not depend on the particular parametrization of the path followed by the light ray, so that we can restrict attention to parametrized curves $q(t)$ where the parameter t varies over a fixed interval $[a, b]$. Then $ds = |\dot{q}|\, dt$, and the functional to be minimized is

$$\mathcal{I}[q] = \int_a^b n(q)\, |\dot{q}|\, dt, \qquad q(a) = A, \qquad q(b) = B, \tag{2.9}$$

with Lagrangian $L = n(q)\, |\dot{q}|$. The Euler-Lagrange equations (2.3) are thus

$$-\frac{d}{dt} \frac{n(q)\, \dot{q}}{|\dot{q}|} + |\dot{q}|\, \frac{\partial n}{\partial q} = 0.$$

It can be seen that the solutions to this equation encode all the usual laws of geometric optics.

We cannot apply the usual Hamiltonian theory to this variational problem since the Lagrangian is degenerate: $\det(\partial^2 L/\partial \dot{q}_i\, \partial \dot{q}_j) \equiv 0$. This is in essence because the solutions are just the curves followed by the light rays, and hence can be reparametrized without affecting whether or not they are solutions to the Euler-Lagrange equations. Indeed, this is a special case of Noether's Second Theorem, [**42**], which states that a Lagrangian is everywhere degenerate if and only if it admits an infinite-dimensional symmetry group. In this case the symmetry group is that of reparametrizations $t \longmapsto \varphi(t)$.

To proceed, we must remove this degeneracy. The easiest way is to assume that the curve is given by the graph of a function, and use one of the coordinates q_i as the parameter. For instance, if we specialize to a planar medium, so $q = (x, y)$, and suppose that the path is given as the graph of a curve $y = f(x)$, the variational problem (2.9) takes

the form

$$\mathcal{I}[y] = \int_a^b n(x,y)\, ds = \int_a^b n(x,y)\, \sqrt{1+\dot{y}^2}\, dx, \qquad y(a) = \alpha, \qquad y(b) = \beta.$$

Now the horizontal coordinate x plays the role of time, and the endpoints are $A = (a, \alpha)$, $B = (b, \beta)$. The Euler-Lagrange equation of this variational problem is

$$-\frac{d}{dx}\left(\frac{n(x,y)\,\dot{y}}{\sqrt{1+\dot{y}^2}} \right) + \frac{\partial n}{\partial y}\sqrt{1+\dot{y}^2} = 0,$$

where $\dot{\ }$ means d/dx. To compute the Hamiltonian form of these equations, the Lagrangian is

$$L(x, y, \dot{y}) = n(x,y)\,\sqrt{1+\dot{y}^2},$$

hence

$$p = \frac{\partial L}{\partial \dot{y}} = \frac{n\,\dot{y}}{\sqrt{1+\dot{y}^2}},$$

which can be explicitly inverted:

$$\dot{y} = \frac{p}{\sqrt{n^2 - p^2}}$$

Therefore, the Hamiltonian is

$$H(p, y) = p\,\dot{y} - L = -\sqrt{n^2 - p^2}$$

with canonical equations

$$\dot{p} = -\frac{\partial H}{\partial y} = \frac{n\,\partial n/\partial y}{\sqrt{n^2 - p^2}}, \qquad \dot{y} = \frac{\partial H}{\partial p} = \frac{p}{\sqrt{n^2 - p^2}}.$$

These are the basic equations describing the paths traced by light rays in geometrical optics.

Poisson Brackets

In classical mechanics, an *observable* is just a function $F(p, q)$ of the coordinates in phase space. If $p(t), q(t)$ are a solution to Hamilton's equations (2.1), then an observable F will, in general depend on t. Its rate of change is given by differentiation:

$$\frac{d}{dt}F(p, q) = \sum_{i=1}^n \left(\frac{\partial F}{\partial q_i}\frac{dq_i}{dt} + \frac{\partial F}{\partial p_i}\frac{dp_i}{dt} \right) = \sum_{i=1}^n \left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i} \right).$$

The summed quantity above plays an extremely important role in the theory, and is called the *Poisson bracket* between the functions F and H, denoted

$$\{F, H\} = \sum_{i=1}^n \left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q_i} \right). \tag{2.10}$$

Thus we have shown that the dynamical rate of change of any observable along solutions to Hamilton's equations is given by

$$\frac{dF}{dt} = \{F, H\}. \tag{2.11}$$

In particular, Hamilton's equations themselves, which express the dynamical rate of change of the coordinates (p, q), can be written in Poisson bracket form:

$$\dot{p} = \{p, H\}, \qquad \dot{q} = \{q, H\},$$

which is sometimes abbreviated

$$\dot{u} = \{u, H\}, \qquad u = (p, q).$$

More generally, the dynamical rate of change of a time-dependent function $F(t, p, q)$ is given by

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \{F, H\}. \tag{2.12}$$

The Poisson bracket (2.10) satisfies the following fundamental properties:

(i) *Bilinearity*: $\{aF + bG, H\} = a\{F, H\} + b\{G, H\}, a, b \in \mathbb{R}$.
(ii) *Skew-symmetry*: $\{F, H\} = -\{H, F\}$.
(iii) *Jacobi identity*: $\{\{F, G\}, H\} + \{\{H, F\}, G\} + \{\{G, H\}, F\} = 0$.
(iv) *Leibniz's rule*: $\{FG, H\} = \{F, H\}G + F\{G, H\}$.

Of particular interest are the fundamental Poisson brackets between the coordinate functions (p, q); these are readily seen to be

$$\{p_i, p_j\} = 0, \qquad \{q_i, q_j\} = 0, \qquad \{q_i, p_j\} = \delta_{ij}, \tag{2.13}$$

where δ_{ij} denotes the *Kronecker delta*, which is 1 if $i = j$ and 0 otherwise. The Poisson bracket is also non-degenerate, meaning that the $2n \times 2n$ skew-symmetric matrix J whose entries are the coordinate brackets $J_{ij} = \{u_i, u_j\}$, $u = (p, q)$, is everywhere nonsingular. Indeed, in canonical coordinates,

$$J = \begin{pmatrix} 0 & -\mathrm{I} \\ \mathrm{I} & 0 \end{pmatrix}, \tag{2.14}$$

where I is the $n \times n$ identity matrix, is the *canonical symplectic matrix*. Darboux' Theorem, [**42**], shows that, locally, for any nondegenerate Poisson bracket satisfying identities (i)–(iv), the underlying manifold must be even-dimensional, and there exist local coordinates $u = (p, q)$ in which it takes the canonical form (2.10).

An observable $F(t, p, q)$ is called a *first integral* or *conservation law* of the Hamiltonian system if it is constant along solutions. According to (2.12), this requires that

$$\frac{\partial F}{\partial t} + \{F, H\} = 0. \tag{2.15}$$

In particular, a time-independent function $F(p, q)$ is a first integral if and only if

$$\{F, H\} = 0. \tag{2.16}$$

Note that by skew symmetry, $\{H, H\} = 0$ for any function H, so if the Hamiltonian function $H(p, q)$ is time-independent, then it is automatically a first integral. This is the law of *Conservation of Energy*.

Also, if H and F are both time-independent, and F is an integral for the Hamiltonian flow determined by H, then H is an integral for the Hamiltonian flow determined by F. By the Jacobi identity, if F and G are first integrals, then so is their Poisson bracket $\{F, G\}$. For instance, if two components of angular momentum, e.g.,

$$p_1 q_2 - p_2 q_1, \qquad p_3 q_1 - p_1 q_3,$$

are conserved, then so is the third since

$$\{p_1 q_2 - p_2 q_1, p_3 q_1 - p_1 q_3\} = p_2 q_3 - p_3 q_2.$$

An important (but trivial) observation is the following: if H does not depend on the coordinate q_1, then the corresponding momentum p_1 is a first integral. In this case, we can reduce the order of the system by 2, only looking at p_i, q_i for $i \geq 2$, and integrating the first equation for q_1 by quadrature. This holds in general: any first integral allows us to reduce the order of the Hamiltonian system by 2; see [**42**] for details. Moreover, any other commuting first integral remains a first integral for the reduced system. If a system posses n functionally independent commuting first integrals:

$$\{F_i, F_j\} = 0, \qquad i, j = 1, \ldots, n,$$

then it is called *completely integrable*, and can (at least in principle) be solved explicitly.

Symmetries and First Integrals

We identify an autonomous system of ordinary differential equations

$$\dot{u}_i = \varphi_i(u), \qquad i = 1, \ldots, n, \tag{2.17}$$

with a vector field

$$\mathbf{v} = \sum_{i=1}^{n} \varphi_i(u) \frac{\partial}{\partial u_i},$$

whose flow is prescribed by the solutions to the system (2.17). If the flow is global (local), at each time t, it defines a (local) diffeomorphism of $\mathbb{R}^n$.

In general, we say a (local) diffeomorphism is a *symmetry* of the system of ordinary differential equations (2.17) if it maps solutions to solutions (where defined). A vector field $\mathbf{w}$ is an *infinitesimal symmetry* if its flow determines a one-parameter group of symmetries. It is not hard to see that, for autonomous systems, $\mathbf{w}$ determines an infinitesimal symmetry of the system of ordinary differential equations for the flow of $\mathbf{v}$ if and only if the two vector fields commute: $[\mathbf{v}, \mathbf{w}] = 0$.

For an autonomous Hamiltonian system (2.1), we define the *Hamiltonian vector field* associated with the Hamiltonian function $H(p,q)$ to be

$$\mathbf{v}_H = \sum_{i=1}^{n} \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i}. \tag{2.18}$$

Note that

$$\mathbf{v}_H(F) = \{F, H\} = -\{H, F\}.$$

We can therefore interpret the Poisson bracket as indicating the infinitesimal change in the function or observable F under the flow induced by the Hamiltonian function H.

The following lemma is crucial, relating Poisson brackets and Lie brackets of Hamiltonian vector fields:

Lemma 2.3. *If F and H are smooth functions with Poisson bracket $G = \{F, H\}$, then the corresponding Hamiltonian vector fields are related according to*

$$\mathbf{v}_G = [\mathbf{v}_H, \mathbf{v}_F] = -[\mathbf{v}_F, \mathbf{v}_H]$$

Proof: Let K be any other function. Then, by the Jacobi identity

$$\mathbf{v}_G(K) = \{K, G\} = \{K, \{F, H\}\} = -\{H, \{K, F\}\} - \{F, \{H, K\}\}$$
$$= \{H, \{F, K\}\} - \{F, \{H, K\}\} = \mathbf{v}_H(\mathbf{v}_F(K)) - \mathbf{v}_F(\mathbf{v}_H(K)). = [\mathbf{v}_H, \mathbf{v}_F](K)$$

Since this holds for all K it is true in general. *Q.E.D.*

Noether's Theorem 2.4. (*Hamiltonian version*) *There is a one-to-one correspondence between Hamiltonian symmetries of a Hamiltonian system, and (time dependent) conservation laws (first integrals).*

Proof: Let $F(p, q)$ be a time-independent first integral. Then, by (2.16), $\{F, H\} = 0$ and hence Lemma 2.3 implies that $\mathbf{v}_F$ and $\mathbf{v}_H$ commute, $[\mathbf{v}_H, \mathbf{v}_F] = 0$, and so $\mathbf{v}_F$ is a symmetry of the Hamiltonian system. Conversely, assuming $\mathbf{v}_F$ is a symmetry of the Hamiltonian system, we have $\mathbf{v}_G = 0$, where $G = \{F, H\}$. This doesn't quite mean that $G = 0$ (unfortunately), but rather that $G = g(t)$ is a function of t alone. However, one can then replace F by $F - \int g\, dt$ to get a time-dependent first integral. The general case is left to the reader. *Q.E.D.*

For example:

Translational symmetry $\longmapsto$ conservation of linear momentum

This is just the observation that when the Hamiltonian doesn't depend on one of the coordinates q_i, then the associated momentum p_i is conserved.

Time translational symmetry $\longmapsto$ conservation of energy

This is the observation that if the Hamiltonian does not depend on t, then it is itself conserved.

Rotational symmetry $\longmapsto$ conservation of angular momentum

The last one is left to the reader to precisely formulate.

Hamilton–Jacobi Theory

There is an intimate connection between systems of first order ordinary differential equations and first order partial differential equations. The two subjects are, in essence, equivalent, and a complete solution to one gives the complete solution to the other. This relationship, going back to Hamilton, Jacobi, and others in the first half of the nineteenth century, lies at the heart of wave/particle duality and the interconnections between classical and quantum mechanics.

We begin with a first order variational problem $\int_{t_0}^{t_1} L(t, q, \dot{q}) \, dt$. Define the *action function* to be

$$S(q, t) = \int_{\gamma} L(t, q, \dot{q}) \, dt, \tag{2.19}$$

where γ is an extremal connecting some fixed initial point (q_0, t_0) to the variable point (q, t). (We suppress this dependence of S on the initial point.) To prove that S is a smooth function defined in a neighborhood of the initial point, we need to know that there is a "central field of extremals" starting at (q_0, t_0) and not intersecting in the neighborhood. This can be proved under our prevailing regularity assumptions using standard existence and uniqueness theorems for ordinary differential equations.

We begin with a key lemma which computes the derivatives of the action function.

Lemma 2.5. *The differential of the action function is given by*

$$dS = p \, dq - H \, dt. \tag{2.20}$$

Remark: The differential one-form on the right hand side of (2.20) is known as the *Cartan form* and serves to define *Hilbert's invariant integral*.

Proof: This is essentially the formula for the general variation of a functional. Suppose we are varying both the function $q(t)$ and the endpoints for the functional

$$\mathcal{I}[q, a, b] = \int_a^b L(t, q, \dot{q}) \, dt, \qquad q(a) = a, \qquad q(b) = b.$$

Let $h(t)$ be a perturbation, but now we also allow the endpoints to vary, as well as the imposed boundary conditions. We compute the associated infinitesimal variation in the integral:

$$\delta I = \frac{d}{d\varepsilon} \bigg|_{\varepsilon = 0} \int_{a + \varepsilon \, \delta a}^{b + \varepsilon \, \delta b} L(t, q + \varepsilon \, h, \dot{q} + \varepsilon, \dot{h}) \, dt = L(t, q, \dot{q}) \bigg|_{t=a}^{b} + \int_a^b \left[\frac{\partial L}{\partial q} \, h(t) + \frac{\partial L}{\partial \dot{q}} \, \dot{h} \right] dt$$

$$= \left(L + \frac{\partial L}{\partial \dot{q}} \, h \right) \bigg|_{t=a}^{b} + \int_a^b \left[\frac{\partial L}{\partial q} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right] h(t) \, dt. \tag{2.21}$$

Now, we have to be a little careful. It we vary the point (t, q) to $(t + \varepsilon \, \delta t, q + \varepsilon \, \delta q)$, the corresponding variation h will be given by

$$q + \varepsilon \, \delta q = q(t + \varepsilon \, \delta t) + \varepsilon \, h(t + \varepsilon \, \delta t) = q(t) + \varepsilon \big[\dot{q}(t) \, \delta t + h(t) \big].$$

Therefore, at the original (unvaried point) t, we have

$$h(t) = \delta q - \dot{q}(t)\,\delta t.$$

Substituting this into the variational formula (2.21), we find that

$$\delta \mathcal{I} = \left[\left(L - \dot{q}\,\frac{\partial L}{\partial \dot{q}}\right)\delta t + \frac{\partial L}{\partial \dot{q}}\,\delta q\right]\Bigg|_{t=a}^{b} + \int_a^b \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right] h(t)\,dt$$

$$= \left[(L - \dot{q}p)\,\delta t + p\,\delta q\right]\Big|_{t=a}^{b} + \int_a^b \left[\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right] h(t)\,dt.$$

since $p = \partial L/\partial \dot{q}$. In particular, if q is a solution to the Euler-Lagrange equations (2.3), then

$$\delta I = \left[(L - \dot{q}p)\,\delta t + p\,\delta q\right]\Big|_{t=a}^{b} = \left(p\,\delta q - H\,\delta t\right)\Big|_{t=a}^{b}.$$

If we only vary one endpoint, formula (2.20) follows. $\hspace{2cm}$ Q.E.D.

The most important consequence of this result is the following.

Theorem 2.6. *The action function is a solution to the* Hamilton–Jacobi *equation:*

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial q}, q, t\right) = 0, \tag{2.22}$$

where $H(p, q, t)$ is the Hamiltonian associated with the variational problem.

Proof: This follows at once from Lemma 2.5, which implies $\partial S/\partial t = -H(p, q, t)$ and $\partial S/\partial q = p$. $\hspace{5cm}$ Q.E.D.

Example 2.7. Consider the special case of Newton's equations for the motion of a mass in a central gravitational force field. The Lagrangian is

$$L(q, \dot{q}) = \frac{m\,|\dot{q}|^2}{2} - V(|q|) = \frac{|p|^2}{2m} - V(|q|),$$

where $p = m\,\dot{q}$. The Hamiltonian is

$$H(p, q) = p\,\dot{q} - L(q, \dot{q}) = \frac{|p|^2}{2m} + V(|q|).$$

Thus, the Hamilton–Jacobi equation (2.22) takes the form

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left|\frac{\partial S}{\partial q}\right|^2 + V(|q|) = 0. \tag{2.23}$$

In spherical coordinates (r, φ, θ), this becomes

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left[\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \varphi}\right)^2 + \frac{1}{r^2 \sin^2 \varphi}\left(\frac{\partial S}{\partial \theta}\right)^2\right] + V(r) = 0. \tag{2.24}$$

Remark: We will use the mathematical convention for spherical coordinates, cf. [**43**], where $-\pi < \theta \leq \pi$ is the *azimuthal angle* or *longitude*, while $0 \leq \varphi \leq \pi$ is the *zenith angle* or *co-latitude*, whereby

$$x = r \sin \varphi \, \cos \theta, \qquad y = r \sin \varphi \, \sin \theta, \qquad z = r \cos \varphi. \qquad (2.25)$$

In many books, particularly those in *physics*, the roles of φ and θ are *reversed*, leading to much confusion when one is perusing the literature.

Example 2.8. In geometric optics, as we have seen, the Hamiltonian is

$$H(p, q) = -\sqrt{n^2 - p^2}$$

with x playing the role of t. The Hamilton–Jacobi equation becomes

$$\frac{\partial S}{\partial x} - \sqrt{n^2 - \left(\frac{\partial S}{\partial y}\right)^2} = 0.$$

which is equivalent to the *eikonal equation* describing optical wave front sets:

$$\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2 = n(x, y)^2.$$

More generally, in n-dimensional optics, the Hamilton–Jacobi equation again coincides with the eikonal equation

$$|\nabla S|^2 = n^2.$$

The solutions to the Hamilton–Jacobi equation are directly related to the solutions to Hamilton's equations. In fact, they are in a sense equivalent differential equations, and knowing the complete solution to one allows one to determine the complete solution to the other. There are several approaches to this; here we invoke the method of characteristics. We will show that the solutions to Hamilton's equations are just the characteristic curves for the Hamilton–Jacobi equation. The physical motivation for this comes from geometrical optics. In this case, the action function S determines the phase of an oscillatory light wave. Its level sets $\{S = c\}$ are the wave front sets where the light waves have constant phase. The characteristics are just the paths followed by the light rays or the photons. This principle is at the heart of Hamilton's "optical-mechanical" analogy, and, ultimately, the basis of the wave/particle duality in quantum mechanics.

Consider a first order partial differential equation

$$F(x, u, \nabla u) = 0, \qquad x = (x_1, \ldots, x_n) \qquad (2.26)$$

where $F(x, u, p)$ is a smooth function on the $(2n+1)$-dimensional space whose coordinates are the independent variables $x = (x_1, \ldots, x_n)$, the dependent variable u, and the gradient coordinates $p = (p_1, \ldots, p_n)$, with p_i representing the derivative $\partial u / \partial x_i$.

The most basic problem associated with such an equation is the *Cauchy problem*, meaning the *initial value problem* in which one specifies the value f of the function u on an $(n-1)$-dimensional submanifold (hypersurface) S of the base space $\mathbb{R}^n$. To solve the

Cauchy problem (classically) one is required to find a solution u with the given initial values: $u|_S = f$. For instance, suppose we adopt the particular coordinates $(t, y) = (t, y_1, \ldots, y_{n-1})$ in which the initial hypersurface is the flat hyperplane $S = \{ (t, y) \,|\, t = 0 \}$. In this case the equation takes the form[†]

$$F(t, y, u, u_t, u_y) = 0. \tag{2.27}$$

The Cauchy data is specified on the hyperplane S by

$$u(0, y) = f(y). \tag{2.28}$$

Suppose we can solve this equation for the normal derivative u_t, placing the equation in *Cauchy–Kovalevskaya form*

$$u_t = G(y, t, u, u_y). \tag{2.29}$$

According to the Implicit Function Theorem, this is possible (locally) provided the partial derivative $\partial F / \partial u_t \neq 0$ does not vanish at a point (t, y, u, u_t, u_y). In this case, the initial hyperplane S is called *non-characteristic*. The Cauchy-Kovalevskaya Existence Theorem, [**17**], shows that, for analytic G, the above Cauchy problem (2.28–29) has a unique solution.

More generally, a hypersurface $S \subset \mathbb{R}^n$ is called *non-characteristic* if the corresponding Cauchy problem

$$u|_S = f$$

is similarly well-posed. Let us represent

$$S = \{ x \,|\, h(x) = 0 \}$$

as the zero locus of a smooth scalar-valued function h with $\nabla h \neq 0$ on S. Assuming without loss of generality that $\partial h / \partial x_n \neq 0$ we can locally introduce new coordinates

$$t = h(x), \qquad y_1 = x_1, \qquad \ldots \qquad y_{n-1} = x_{n-1},$$

to flatten out S to be the hyperplane $\{ t = 0 \}$. Then

$$\frac{\partial u}{\partial x_j} = \xi_j \frac{\partial u}{\partial t} + \frac{\partial u}{\partial y_j}, \qquad 1 \leq j \leq n - 1, \qquad \frac{\partial u}{\partial x_n} = \xi_n \frac{\partial u}{\partial t}, \qquad \text{where} \qquad \xi_j = \frac{\partial h}{\partial x_j}.$$

The Implicit Function Theorem requires that, to be able to solve smoothly for the "normal" derivative u_t, we must have

$$0 \neq \frac{\partial F}{\partial u_t} = \sum_{j=1}^{n} \xi_j \frac{\partial F}{\partial p_j}.$$

All of this discussion serves as motivation for the following crucial definition.

[†] We now make use of subscript notation to denote partial derivatives, so that u_t represents $\partial u / \partial t$, while u_y represents the y gradient $\nabla_y u$ with entries $u_{y_i} = \partial u / \partial y_i$.

Definition 2.9. An n-tuple of numbers $\xi = (\xi_1, \ldots, \xi_n)$ determines a *characteristic direction* for the partial differential equation $F(x, u, p) = 0$ at the point (x, u, p) if

$$\xi \cdot F_p(x, u, p) = 0.$$

Then, very loosely speaking, a hypersurface $S = \{h(x) = 0\}$ is *characteristic* if its *normal* $\xi = \nabla h$ determines a characteristic direction at each point. In point of fact, the components of ξ transform like a differential one-form $dh = \sum \xi_j dx_j$, i.e., a covariant 1-tensor, and should be treated as such.

Away from singular points where the derivative F_p vanishes, the characteristic directions span an $(n-1)$-dimensional subspace of (the cotangent space to) $\mathbb{R}^n$. Throughout this section, we will *always* assume that $F_p \neq 0$ so as to avoid singularities of the equation. The orthogonal complement to this characteristic subspace will be a one-dimensional subspace of (the tangent space to) $\mathbb{R}^n$. Thus we will call a tangent vector $\mathbf{v} = (v_1, \ldots, v_n)$ at a point $x \in \mathbb{R}^n$ a *characteristic vector* for the point (x, u, p) sitting over x if it is orthogonal to all the characteristic directions ξ at this point, i.e., $\mathbf{v} \cdot \xi = 0$ for all ξ such that $\xi \cdot F_p = 0$. Clearly this is true if and only if $\mathbf{v}$ is parallel to F_p, i.e., $\mathbf{v} = \lambda F_p$ for some scalar λ. In particular, except in the special case of a linear equation, we need to know the values of u and p in order to specify the characteristic vectors at a point x. A hypersurface S is characteristic if and only if the corresponding characteristic vector $\mathbf{v}$ is contained its tangent space at each point.

An alternative, but equivalent definition of a characteristic direction is a direction in which the derivative of some solution admits a possible discontinuity. Note that, if u is a continuous solution to the equation that is C^1 except on the hypersurface $S = \{t = 0\}$, where the normal derivative u_t has a discontinuity, then S must be characteristic, since if we can solve for u_t as above, then it is necessarily continuous, as all the functions on the right hand side are continuous. Both definitions of characteristics extend to higher order equations, although they can typically no longer be used to derive the general solution.

The astute reader will have realized that the above discussion is not very precise. Namely, to determine whether or not a hypersurface is characteristic we need to check whether or not its normal, which depends on x, is a characteristic direction, but, except in the special case of linear equations, the condition specifying the characteristic directions depends on the values of x, u, p, and we have yet to specify u and p. These will come from the solution $u = f(x)$ of the equation, but the reasoning is starting to get a little circular. It helps at this point to step back, and describe what we mean by the Cauchy problem a little more precisely. First we introduce the base (physical) space $X = \mathbb{R}^n$, which has coordinates $x = (x_1, \ldots, x_n)$. Sitting over X is the $(2n+1)$-dimensional space J, which has coordinates $(x, u, p) = (x_1, \ldots, x_n, u, p_1, \ldots, p_n)$ representing the dependent variable u and its first order derivatives p. The space J is often referred to as the *first jet space*, and denoted J^1. The standard projection $\pi: J \longrightarrow X$ maps $\pi(x, u, p) = x$.

Consider a smooth function $u = f(x)$ defined on a domain $D \subset X$, so f is a function from D to $\mathbb{R}$. By the *prolongation* of f, we mean the function $f^{(1)}: D \longrightarrow J$, given by taking the derivatives, so $u = f(x)$, $p = \nabla f(x)$ for $x \in D$. The graph of f determines an n-dimensional submanifold of $\mathbb{R}^{n+1} = X \times \mathbb{R}$. Similarly, the graph of the prolongation

$f^{(1)}$ determines an n-dimensional submanifold $\Gamma_f \subset J$, given by

$$\Gamma_f = \{\, (\, x, f(x), \nabla f(x)\,) \mid x \in D \,\}.$$

However, not every n-dimensional submanifold $N \subset J$ is the graph of a function; the derivative coordinates p must match up properly. The easiest way to express this condition is to use the *contact one-form*

$$\theta = du - p\, dx = du - \sum_{i=1}^n p_i\, dx_i. \tag{2.30}$$

Note that if we substitute $u = f(x)$, $p = \nabla f(x)$, into θ, then the resulting one-form vanishes identically. In other words, the contact one-form θ vanishes when restricted to the prolonged graph of a function. Subject to a certain transversality condition, the converse to this also holds. We formalize this condition into a general definition.

Definition 2.10. A submanifold $N \subset J$ is called *Lagrangian* if the contact one form (2.30) vanishes on it: $\theta|_N = 0$.

As for the transversality condition, consider the volume form

$$\Omega = dx_1 \wedge dx_2 \wedge \ldots \wedge dx_n, \tag{2.31}$$

which is also defined on the jet space J. Note that Ω does not vanish on the (prolonged) graph of any function $u = f(x)$. Conversely, if we have a parametrized n-dimensional submanifold $N = \{\, (x(s_1, \ldots, s_n), u(s_1, \ldots, s_n), p(s_1, \ldots, s_n)) \,\} \subset J$, then, by the Inverse Function Theorem, we can locally express u as a function of the x's provided the Jacobian determinant $\det(\partial x_i / \partial s_j) \neq 0$. This is the same as saying that the volume form Ω does not vanish on N.

Proposition 2.11. *Suppose N is an n-dimensional Lagrangian submanifold of J such that $\Omega \neq 0$ on N. Then, locally near each point x corresponding to some $(x, u, p) \in N$, we can express $N = \Gamma_f$ as the prolonged graph of a function $u = f(x)$.*

Next, consider the standard Cauchy problem $u(0, y) = f(y)$ on an initial hyperplane $S = \{\, (t, y) \mid t = 0 \,\}$ in the (t, y) coordinates. If the problem is non-characteristic, then this Cauchy data will determine initial conditions not only for u on the hyperplane S but also the derivatives $p = (u_t, u_y)$ on S as follows. The tangential y-derivatives come directly from $f(y)$ itself: $u_{y_i} = \partial f / \partial y_i$. As for the t-derivative u_t, we assume that we have solved the equation for $u_t = G(t, y, u, u_y)$. Then, on the initial hypersurface S, the normal derivative is given by $u_t = G(0, y, f(y), f_y(y))$. Therefore, the standard Cauchy data parametrizes an $(n-1)$-dimensional submanifold of the jet space J, namely

$$N_0 = \left\{ \left(0, y, f(y), G\left(y, 0, f(y), \frac{\partial f}{\partial y}(y) \right), \frac{\partial f}{\partial y}(y) \right) \right\}.$$

The solution to this Cauchy problem will then be a function $u = f(t, y)$ such that $f(0, y) = f(y)$. It is easy to see that this is the same as requiring that the prolonged graph Γ_f of

$f(t, y)$ contain the "Cauchy submanifold" N_0, i.e., $\Gamma_f \supset N_0$. Similar properties hold in the general coordinates x.

Thus, we are tempted to define the Cauchy data of the partial differential equation as an $(n-1)$-dimensional submanifold of the jet space J. However, as with the prolonged graph of a function, not every such submanifold can serve as valid Cauchy data. First, it must itself satisfy the equation, i.e., consist of points (x, u, p) such that $F(x, u, p) = 0$. Secondly, in order that the derivatives match up, the Cauchy submanifold must itself be Lagrangian; otherwise the tangential derivatives will not be correct. Thirdly, it must project back down to an $(n-1)$-dimensional submanifold $S = \pi(N_0)$ of the base space X.

This inspires the following important definition.

Definition 2.12. Consider a first order partial differential equation $F(x, u, p) = 0$ defined by the vanishing of a smooth real-valued function on the jet space $J = \mathbb{R}^{2n+1}$. A *Cauchy submanifold* is an $(n-1)$-dimensional Lagrangian submanifold $N_0 \subset J$ which satisfies the equation

$$N_0 \subset \{ (x, u, p) \mid F(x, u, p) = 0 \},$$

and whose projection $S = \pi(N_0)$ is an $(n-1)$-dimensional submanifold of the base space X. A *solution* to the Cauchy problem defined by N_0 will be a smooth solution $u = f(x)$ of the partial differential equation whose prolonged graph contains the Cauchy submanifold: $\Gamma_f \supset N_0$. The Cauchy submanifold N_0 is called *non-characteristic* if for each point $(x, u, p) \in N_0$, the characteristic vector $\mathbf{v} = F_p$ is *not* tangent to its projection S at the corresponding point $x = \pi(x, u, p)$.

It is often convenient to replace the projection condition by a local transversality condition. In the non-characteristic case, this can be written as

$$\mathbf{v} \lrcorner \, \Omega \neq 0 \qquad \text{on} \qquad S = \pi(N_0).$$

Here $\mathbf{v} = F_p$ is the characteristic vector, Ω is the volume n-form on $\mathbb{R}^n$, and $\lrcorner$ denotes *interior product*, so that $\mathbf{v} \lrcorner \, \Omega$ is the unique $(n-1)$-form that satisfies

$$\langle \mathbf{v} \lrcorner \, \Omega, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \rangle = \langle \Omega, \mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_{n-1} \rangle \tag{2.32}$$

for all tangent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$.

If N_0 is parametrized explicitly by $(x(s_1, \ldots, s_{n-1}), u(s_1, \ldots, s_{n-1}), p(s_1, \ldots, s_{n-1}))$, then the conditions that N_0 be a non-characteristic Cauchy submanifold have the following equivalent coordinate versions:

(i) satisfies the equation: $F(x(s_1, \ldots, s_{n-1}), u(s_1, \ldots, s_{n-1}), p(s_1, \ldots, s_{n-1})) = 0$.

(ii) Lagrangian: $\dfrac{\partial u}{\partial s_j} - \displaystyle\sum_{i=1}^{n} p_i \dfrac{\partial \xi}{\partial s_j} = 0, \quad j = 1, \ldots, n-1$.

(iii) non-characteristic: $\det \left(\dfrac{\partial F}{\partial p_i}, \dfrac{\partial \xi}{\partial s_j} \right) \neq 0$.

The Cauchy–Kovalevskaya Theorem now guarantees the unique solution to the Cauchy problem associated with a non-characteristic Cauchy submanifold, provided the equation

and initial data are analytic. Actually, we will see how to solve this problem under much weaker hypotheses.

Let $u = f(x)$ be any C^2 solution to the equation $F = 0$ defined on a domain $D \subset X$. At each point $x \in D$, the solution will determine (up to multiple) a characteristic vector $\mathbf{v} = \lambda F_p(x, f(x), f_x(x))$. A parametrized curve $x(s)$ whose non-zero tangent vector is everywhere a characteristic vector for the given solution u will be called a *characteristic curve* for the solution. This requires that $\dot{x}(s)$ be proportional to F_p at the point:

$$\frac{dx}{ds} = \lambda \frac{\partial F}{\partial p}.$$

Now, except in the case of linear equations, the characteristic vectors depend not only on the base point x, but also on the value of u and its derivatives p at x. This suggests that, in order to determine a characteristic curve, we not only look at the base curve $x(s)$, but also its "prolongation" to the jet space J. There will be a unique curve contained in the prolonged graph Γ_f of the solution sitting over the base curve $x(s)$, namely the curve $(x(s), u(s), p(s))$, where

$$u(s) = f(x(s)), \qquad p(s) = \frac{\partial f}{\partial x}(x(s)).$$

We can use the chain rule to compute how the u and p components of the prolonged curve depend on s. Thus,

$$\frac{du}{ds} = \sum_{i=1}^{n} \frac{\partial u}{\partial x_i} \frac{dx_i}{ds} = \lambda \sum_{i=1}^{n} p_i \frac{\partial F}{\partial p_i} = \lambda p \frac{\partial F}{\partial p}.$$

and also,

$$\frac{dp_i}{ds} = \sum_k \frac{\partial p_i}{\partial x_k} \frac{dx_k}{ds} = \lambda \sum_k \frac{\partial^2 u}{\partial x_i \partial x_k} \frac{\partial F}{\partial p_k}.$$

At this stage it appears that we also need to know how the second derivatives of the solution behave. However, u is assumed to be a solution, so it also satisfies

$$0 = \frac{\partial}{\partial x_i} F(x, u, u_x) = \frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial u} \frac{\partial u}{\partial x_i} + \sum_k \frac{\partial^2 u}{\partial x_i \partial x_k} \frac{\partial F}{\partial p_k}.$$

Comparing these two equations, we see that

$$\frac{dp_i}{ds} = -\lambda \left(\frac{\partial F}{\partial x_i} + p_i \frac{\partial F}{\partial u} \right).$$

Finally, note that we can absorb the proportionality factor λ into the parameter s by reparametrizing the curve, so $\lambda = 1$ without loss of generality. We are thus left with a system of first order ordinary differential equations for the components $(x(s), u(s), p(s))$ which do not refer any longer to the particular solution $u = f(x)$.

Definition 2.13. A *characteristic curve* is a solution to the characteristic system of ordinary differential equations:

$$\frac{dx}{ds} = \frac{\partial F}{\partial p}, \qquad \frac{du}{ds} = p\,\frac{\partial F}{\partial p}, \qquad \frac{dp}{ds} = -\left(\frac{\partial F}{\partial x} + p\,\frac{\partial F}{\partial u}\right). \qquad (2.33)$$

In general, given a point $(x_0, u_0, p_0) \in J$, we will denote by $(x(s), u(s), p(s))$ the characteristic curve passing through it, whereby (x_0, u_0, p_0) are the initial conditions for (2.33) at $s = 0$. Note that, by standard existence and uniqueness results for ordinary differential equations, this curve is uniquely defined for s sufficiently small.

Thus, we have demonstrated that any characteristic curve associated with a solution u of the partial differential equation is a solution to the characteristic system of ordinary differential equations (2.33). Since the characteristic curves are uniquely determined by their initial conditions, we deduce that every solution to our partial differential equation is swept out by an $(n-1)$–parameter family of characteristic curves, parametrized by the initial values of u and p on the $(n-1)$-dimensional non-characteristic Cauchy surface.

If F doesn't depend on u, then the characteristic equations (2.33) are essentially the same as Hamilton's equations for a time independent Hamiltonian $F = H(p, q)$ (with $x = q$) since u can be determined from x, p by a single quadrature. For the *time-dependent Hamilton–Jacobi equation*

$$F(t, q, S, S_t, S_q) \equiv \frac{\partial S}{\partial t} + H\left(t, \frac{\partial S}{\partial q}, q\right) = 0, \qquad (2.34)$$

we replace S by u to find the corresponding equations for characteristic curves

$$\frac{dt}{ds} = 1, \qquad \frac{dq}{ds} = \frac{\partial H}{\partial p}, \qquad \frac{d\pi}{ds} = -\frac{\partial H}{\partial t}, \qquad \frac{dp}{ds} = -\frac{\partial H}{\partial q}, \qquad \frac{du}{ds} = \pi + p\,\frac{\partial H}{\partial p}, \quad (2.35)$$

where π represents S_t. Thus $t = s + c$, and, after we solve the Hamiltonian system for p, q, we can recover the complete expression for the characteristic curves by quadrature.

Now we prove the converse — that every solution $u = f(x)$ to the partial differential equation can be constructed from the characteristic curves.

Proposition 2.14. *Let $u = f(x)$ be a solution to the first order partial differential equation (2.26). Let $x_0,\ u_0 = f(x_0),\ p_0 = f_x(x_0)$ be a point belonging to the (prolonged) graph of f. Let $(x(s), u(s), p(s))$ be the solution to the characteristic system (2.33) with initial conditions (x_0, u_0, p_0). Then this curve is a characteristic curve for the solution, i.e., it is contained in the prolonged graph of f for all s.*

Thus if we are given non-characteristic Cauchy data, we will solve the characteristic system (2.33) for each associated initial point, and then reconstruct the solution by piecing together characteristic curves. Now, given the initial manifold N_0 corresponding to the Cauchy data, let

$$N_s = \{\,(x(s), u(s), p(s)) \mid (x_0, u_0, p_0) \in N_0\,\}.$$

be the image under the characteristic flow at "time" s, provided this is defined. For N_0 bounded, and s sufficiently small, N_s is defined and smooth. Let N^* be the union of

all the $(n-1)$-dimensional submanifolds N_s for $|s| \leq s^*$, so N^* is an n-dimensional submanifold of the jet space. Then, as long as N_0 is a valid Cauchy submanifold and s^* is sufficiently small, the family of solutions N^* will (locally) be the prolonged graph of a solution $u = f(x)$ to the partial differential equation.

Theorem 2.15. *Consider the Cauchy problem for the first order partial differential equation $F(x, u, p) = 0$ determined by an $(n-1)$-dimensional Cauchy submanifold N_0 of the jet space J. Let $N^* = \bigcup N_s$ be the n-dimensional submanifold with initial data N_0 swept out by the characteristic flow. Then, locally near N_0, the submanifold N^* is the prolonged graph of a solution $u = f(x)$, that is, $N^* = \Gamma_f$ near N_0.*

Proof: There are two basic steps. First, we prove that N^* is a solution to the partial differential equation, i.e., $N^* \subset \{F = 0\}$. Second, we show, using the previous proposition, that N^* is locally the prolonged graph of some function. For the first step, it suffices to note that $F(x, u, p)$ itself is a first integral of the characteristic system of ordinary differential equations. Indeed

$$\frac{d}{ds}F = \frac{\partial F}{\partial x}\frac{dx}{ds} + \frac{\partial F}{\partial u}\frac{du}{ds} + \frac{\partial F}{\partial p}\frac{dp}{ds} = F_x F_p + F_u(pF_p) + F_p(-F_x - pF_u) = 0.$$

We conclude that F is constant along solutions. In particular, if $F = 0$ initially, it remains zero for all s. Therefore, if the initial data N_0 satisfies the equation, so does each N_s, and hence all of N^*.

As for the second step, we need to prove that N^* is Lagrangian, and, at least near N_0, satisfies the transversality condition. The Lagrangian condition is equivalent to the fact that the coordinates p are indeed the derivatives of u. We compute the Lie derivative of the contact one-form:

$$\frac{d\theta}{ds} = d\left(\frac{du}{ds}\right) - \frac{dp}{ds}\,dx - p\,d\left(\frac{dx}{ds}\right) = d(pF_p) - p\,dF_p + (F_x + pF_u)\,dx$$

$$= F_p\,dp + F_x\,dx + pF_u\,dx = dF - F_u(du - p\,dx) = dF - F_u\,\theta.$$

Now, we already know that $F = 0$, since we assumed that it was zero initially. Thus, $dF = 0$, and hence $d\theta/ds = -F_u\,\theta$. Therefore, if θ vanishes on N_0 it also vanishes on each N_s, and hence on their union N^*. Thus N^* is Lagrangian.

Finally, to prove transversality, the fact that N_0 is not characteristic means that the characteristic vector $\mathbf{v}$ is not tangent to its projection $S = \pi(N_0)$, and so at each point of S we can choose a basis $\mathbf{w}_1, \ldots, \mathbf{w}_{n-1}$ for its tangent space and know that the full set of n tangent vectors $\mathbf{v}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1}$ forms a basis for the tangent space to $X = \mathbb{R}^n$ at the point. To check that the n-form Ω does not vanish on N^*, at least in a neighborhood of N_0, by continuity we need only prove that it does not vanish on N_0 itself. But, by (2.32)

$$\langle \Omega, \mathbf{v}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1} \rangle = \langle \mathbf{v} \lrcorner \Omega, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1} \rangle \neq 0 \quad \text{on} \quad N_0. \qquad\qquad Q.E.D.$$

Note that we cannot, in general, expect the transversality condition $\Omega \neq 0$ to hold everywhere. For instance, in optics, the characteristic curves $(x(s), p(s))$ project to the light rays $x(s)$. But the crossing of light rays is precisely the phenomena associated with

focussing and the formation of caustics. The points where $\Omega = 0$ correspond to points where the submanifold N^* in jet space no longer projects to a single-valued function on physical space, and will thus be associated with such breakdowns of simple physical phenomena. This could lead us into the study of shock waves and caustics, leading into catastrophe theory.

Therefore, the integration of a first order partial differential equation is reduced to the integration of a system of first order ordinary differential equations. Remarkably, the converse also holds: if we know the general solution to a first order partial differential equation, we can recover the solution to the corresponding characteristic system of ordinary differential equations.

First consider one solution $u(x)$ to the first order partial differential equation

$$F(x, u, p) = 0. \tag{2.36}$$

Then the characteristic curves embedded in the graph of u can be found by integrating a first order system of ordinary differential equations, viz.

$$\frac{dx}{ds} = \frac{\partial F}{\partial p}(x, u(x), u_x(x)).$$

Indeed, if $x(s)$ is the solution to this system, and we define

$$u(s) = u(x(s)), \qquad p(s) = u_x(x(s)),$$

then it is easy to see that $(x(s), u(s), p(s))$ solves the full characteristic system. This is the same computation as above:

$$\frac{du}{ds} = \frac{\partial u}{\partial x}\frac{dx}{ds} = p\frac{\partial F}{\partial p}, \qquad \frac{dp}{ds} = \frac{\partial p}{\partial x}\frac{dx}{ds} = \frac{\partial^2 u}{\partial x^2}\frac{\partial F}{\partial p} = -\left(\frac{\partial F}{\partial x} + p\frac{\partial F}{\partial u}\right).$$

For example, consider the Hamilton–Jacobi equation (2.22). If $S(t, q)$ is a solution, then the associated characteristic curves (solutions to Hamilton's equations) are found by integrating the first order system

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}(t, q, S_q(q, t)) \tag{2.37}$$

for the positions, and then substituting

$$p(t) = \frac{\partial S}{\partial q}(t, q(t)).$$

to get the corresponding momenta. The solution $q(t)$ to (2.37) will describe the particle trajectories in physical space.

As another example, if we know a solution $S(q)$ to the *eikonal equation*

$$|\nabla S|^2 = n(q) \tag{2.38}$$

of geometrical optics, then the associated light rays are found by integrating

$$\frac{dq}{dt} = 2\,\nabla S(q),$$

since $F = p^2 - n(q)$, so $F_p = 2p$. We deduce the important fact that the light rays are orthogonal to the wave front sets $\{S = c\}$. In fact, (2.37) can be interpreted as the corresponding "orthogonality" relation for mechanics.

Next, suppose we know lots of solutions to our first order partial differential equation (2.26). Then we should be able to say a lot more about solutions to the corresponding characteristic system. For instance, if we have a one-parameter family of solutions $u(x, \lambda)$, then they all contain characteristic curves in their graphs, so that envelopes of these solutions will be composed of characteristic curves. Recall that the envelope is constructed by solving the equation $\partial u / \partial \lambda = 0$ for λ as a function of x and substituting back into $u(x, \lambda)$. This suggests that we look at the quantity $u_\lambda = \partial u / \partial \lambda$ in detail. Its evolution along characteristics is given by

$$\frac{du_\lambda}{ds} = u_{\lambda x} \frac{dx}{ds} = p_\lambda F_p, \qquad (2.39)$$

where $p(x, \lambda) = u_x(x, \lambda)$. On the other hand, differentiating

$$F(x, u(x, \lambda), p(x, \lambda)) = 0$$

with respect to λ, we deduce that

$$0 = F_u u_\lambda + F_p p_\lambda, \qquad \text{hence} \qquad \frac{du_\lambda}{ds} = -F_u u_\lambda.$$

Note that $u_\lambda = 0$ is a solution to this equation, confirming our earlier deduction about envelopes. In particular, if F does not depend explicitly on u (as is the case in the Hamilton–Jacobi equation or the eikonal equation) then $\mathbf{u}_\lambda$ is a first integral of the characteristic system.

In the more general case, we need a two-parameter family of solutions $u(x, \lambda_1, \lambda_2)$. Then the ratio $u_{\lambda_1} / u_{\lambda_2}$ will be a first integral of the characteristic system, since, by (2.39),

$$\frac{d}{ds} \frac{u_{\lambda_1}}{u_{\lambda_2}} = \frac{u_{\lambda_1 s} u_{\lambda_2} - u_{\lambda_1} u_{\lambda_2 s}}{u_{\lambda_2}^2} = \frac{-F_u u_{\lambda_1} u_{\lambda_2} - u_{\lambda_1}(-F_u u_{\lambda_2})}{u_{\lambda_2}^2} = 0.$$

Note that this includes the case when F doesn't depend on u, since in that situation, if $u(x)$ is any solution, so is $u(x) + \mu$ for any constant μ. Therefore, we can replace our (nontrivial) one-parameter family of solutions $u(x, \lambda)$ by a two-parameter family $u(x, \lambda_1) + \lambda_2$, whereby the ratios reduce to the earlier first integrals. However, the parameters λ_1, λ_2 must enter into $u(x, \lambda_1, \lambda_2)$ in an essentially independent way for this to be of any benefit. For instance, if we use just a one-parameter family of solutions, and try $u(x, \varphi(\lambda_1, \lambda_2))$, then the above first integral is just a function of λ_1, λ_2, and hence of no use.

Now, if we have enough first integrals, then we can solve the characteristic system. This means that we have a family of solutions with enough independent parameters in it. We make the following classical definition.

Definition 2.16. A function $u(x, \lambda)$ depending on n parameters $\lambda = (\lambda_1, \ldots, \lambda_n)$ is called a *complete integral* to the first order partial differential equation

$$F(x, u, p) = 0$$

if for each fixed value of λ it provides a solution, and moreover, that the $n \times (n+1)$ matrix

$$\left(\frac{\partial u}{\partial \lambda_j}, \ \frac{\partial^2 u}{\partial x_i \partial \lambda_j} \right) \tag{2.40}$$

has rank n.

The second condition is to ensure that the parameters λ enter truly independently into the solution. This condition can be seen to be equivalent to the statement that the union of the prolonged graphs of the functions $u(x, \lambda)$ fills out an open subset of the locus

$$\{ (x, u, p) \mid F(x, u, p) = 0 \}$$

defining the equation.

It can be proved that any solution to a first order partial differential equation can be obtained from the complete integral. Essentially, one calculates the envelope to a general $n - 1$ parameter family of solutions by assigning arbitrary functions $\lambda_j = \gamma_j(s_1, \ldots, s_{n-1})$ to the parameters and eliminating the s's from the envelope conditions

$$\sum_{j=1}^{n} \frac{\partial u}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial s_k} = 0, \qquad k = 1, \ldots, n-1.$$

Moreover, the complete integral also gives the solution to the corresponding characteristic system of ordinary differential equations. Note first that to find the general solution to the characteristic system, which is of order $2n + 1$, we need to find solutions depending on $2n + 1$ arbitrary constants. However, two of the constants are trivial. First, the system is autonomous, so we can always specify the initial time t_0 at will; equivalently, we can use one of the variables q_j as the parameter, and rewrite everything in terms of this new variable, whereby we just deduce the unparameterized characteristic curves. Second, the function $F(x, u, p)$ is, as we have seen, always a first integral, so we can fix its value. Indeed, if we are solving the equation $F = 0$, then the value of F is fixed at 0, but the more general equation $F = c$ leads to the same characteristic system, and elucidates the second constant c more directly. We thus need $2n - 1$ additional constants, and these are provided by the parameters $(\lambda_1, \ldots, \lambda_n)$ in the complete integral and the associated first integrals $c_i = u_{\lambda_i}/u_{\lambda_n}$, $i = 1, \ldots, n-1$, deduced above. Note that there are just enough to do this, provided they are "independent".

Theorem 2.17. *Suppose $u(x, \lambda)$ is a complete integral to the partial differential equation $F(x, u, p) = 0$. Assume we are at a point where $u_{\lambda_n} \neq 0$. Then for each value of the $2n - 1$ parameters $(\lambda_1, \ldots, \lambda_n, c_1, \ldots, c_{n-1})$, the equations*

$$\frac{u_{\lambda_i}(x, \lambda)}{u_{\lambda_n}(x, \lambda)} = c_i, \qquad i = 1, \ldots, n-1, \tag{2.41}$$

implicitly determine a curve $\Gamma_{\lambda,c} \subset X$. The complete characteristic curve sitting over $\Gamma_{\lambda,c}$ is then given implicitly by

$$C_{\lambda,c} = \left\{ (x, u(x, \lambda), p(x, \lambda)) \mid x \in \Gamma_{\lambda,c} \right\}.$$

The only thing to prove is that (2.41) really do implicitly determine a curve, which will, by the Implicit Function Theorem, be the case provided the $n \times (n-1)$ Jacobian matrix $(\partial c_i / \partial x_k)$ has rank $n-1$. The entries of this matrix are

$$\frac{\partial}{\partial x_k} \frac{u_{\lambda_i}}{u_{\lambda_n}} = \frac{u_{\lambda_i x_k} u_{\lambda_n} - u_{\lambda_i} u_{\lambda_i x_k}}{u_{\lambda_n}^2}.$$

It is not hard to see that this is the $n \times (n-1)$ submatrix sitting in the matrix obtained from (2.40) by applying the elementary row operation of subtracting $u_{\lambda_i}/u_{\lambda_n}$ times the nth row from the ith row, and then deleting the first column and last row. Since the full matrix (2.40) has rank n, this submatrix must have rank $n-1$. $\qquad$ Q.E.D.

In the case of a Hamilton–Jacobi equation

$$S_t + H(t, S_q, q) = 0,$$

a complete integral has the form

$$S(x, \lambda) + \mu, \qquad \lambda = (\lambda_1, \ldots, \lambda_n),$$

such that the $n \times n$ matrix

$$\left(\frac{\partial^2 u}{\partial q_i \partial \lambda_j} \right)$$

is nonsingular. The corresponding solution to Hamilton's equations is given by

$$c = \frac{\partial S}{\partial \lambda}, \qquad p = \frac{\partial S}{\partial q}.$$

Note that the first equation says that the derivatives of u with respect to the parameters λ are first integrals of Hamilton's equations.

As an application, we can solve the central force field problem from Example 2.7 in this manner. The Hamilton–Jacobi equation (2.24) can be solved by additive separation of variables:

$$S(t, r, \varphi, \theta) = R(r) + \Phi(\varphi) + \alpha \, \theta + \beta \, t.$$

The functions $R(r), \Phi(\varphi)$ satisfy the pair of ordinary differential equations

$$\frac{1}{2m} \left(\frac{dR}{dr} \right)^2 + V(r) + \beta = \gamma \, r^{-2}, \qquad \left(\frac{d\Phi}{d\varphi} \right)^2 + \frac{\alpha^2}{\sin^2 \varphi} + \gamma = 0,$$

where γ is the separation constant. These can be straightforwardly integrated by quadrature, resulting in the required three-parameter solution. One can use this to derive all of Kepler's laws of planetary motion.

3. Geometric Optics and Wave Mechanics.

References: [**10**], [**32**], [**35**], [**47**].

The derivation of the equations for geometric optics from those of wave optics provides the key to Schrödinger's establishment of the basic equation of quantum mechanics, and also the classical limit of quantum mechanical phenomena. Here we outline the derivation of the geometric optics approximation to the equations of wave theory, and then extend these results to obtain the Schrödinger equation.

The Wave Equation

For simplicity, we first consider the scalar *wave equation*

$$\varphi_{tt} - \frac{c^2}{n^2}\,\Delta\varphi = 0 \tag{3.1}$$

in an inhomogeneous medium, where $n(\mathbf{x})$ is the *index of refraction*. In the homogeneous case where n is constant, the solutions to this problem are superpositions of plane waves

$$\varphi = A\,e^{\,\mathrm{i}\,(\,\mathrm{k}\,\cdot\,\mathbf{x} - \omega\,ct)},$$

where the *wave number* (spatial frequency) k and the temporal *frequency* ω are connected by the *dispersion relation*

$$|\,\mathrm{k}\,| = n\,\omega.$$

In the geometrical optics approximation, we consider the case when the wave number is large in comparison to the variation in the refractive index. We begin by restricting our attention to a simple harmonic wave

$$\varphi(t,\mathbf{x}) = u(\mathbf{x})\,e^{\,\mathrm{i}\,\omega\,ct}.$$

This allows us to factor out the t dependence in the wave equation, implying that u satisfies the *Helmholtz equation*

$$\Delta u + n^2\,\omega^2\,u = 0. \tag{3.2}$$

We seek complex solutions of the form

$$u(\mathbf{x}) = A(\mathbf{x},\omega)\,e^{\,\mathrm{i}\,\omega S(\mathbf{x},\omega)}, \tag{3.3}$$

where the amplitude $A(\mathbf{x},\omega)$ and the phase $S(\mathbf{x},\omega)$ are real. Substituting (3.3) into the Helmholtz equation (3.2) yields

$$\Delta u + n^2\omega^2 u = \left[\,\omega^2\left(-|\nabla S|^2 + n^2\right)A + \Delta A + \mathrm{i}\,\omega\left(2\,\nabla S\cdot\nabla A + (\Delta S)\,A\right)\right]e^{\,\mathrm{i}\,\omega S}.$$

Since both A and S are real, we have the system

$$\Delta A + \omega^2\left(-|\nabla S|^2 + n^2\right)A = 0, \qquad 2\,\nabla S\cdot\nabla A + (\Delta S)\,A = 0. \tag{3.4}$$

which is, so far, completely equivalent to the Helmholtz equation.

Now in the high frequency approximation, we introduce asymptotic expansions of the amplitude and phase,

$$A(\mathbf{x}, \omega) \sim A(\mathbf{x}) + \omega^{-1} A_1(\mathbf{x}) + \omega^{-2} A_2(\mathbf{x}) + \cdots,$$
$$S(\mathbf{x}, \omega) \sim S(\mathbf{x}) + \omega^{-1} S_1(\mathbf{x}) + \omega^{-2} S_2(\mathbf{x}) + \cdots,$$

in decreasing powers of the frequency ω. Substituting into the system (3.4), we collect terms involving the same powers of ω. The leading term, in ω^2, occurs in the first equation. Since $A \neq 0$, we deduce that the leading phase $S(\mathbf{x})$ must satisfy

$$|\nabla S|^2 = n^2, \tag{3.5}$$

which is the *eikonal equation* we already encountered in (2.38). It says that the hypersurfaces of constant phase $\{S = c\}$ are the same as the characteristic hypersurfaces for the wave equation. If we interpret S as the action function, then the eikonal equation is the same as the Hamilton–Jacobi equation (2.22) for the geometric optics Hamiltonian. Thus, the phase surfaces propagate along the characteristics, which are just the solutions to Hamilton's equations.

The next term, of order ω, says that the leading amplitude $A(\mathbf{x})$ will satisfy the *transport equation*

$$2 \nabla S \cdot \nabla A + A \Delta S = 0. \tag{3.6}$$

To solve this equation, suppose the curve $(p(s), q(s))$ in phase space determines a characteristic for the associated characteristic system, where

$$F(q, u, p) = p^2 - n(q)^2.$$

Then, if $\Phi(q)$ is any function of position, we have

$$\frac{d\Phi}{ds} = \{ \Phi, F \} = F_p \Phi_q = 2p \, \Phi_q = 2 \nabla S \cdot \nabla \Phi.$$

Therefore, along the characteristics, the transport equation (3.6) reduces to an ordinary differential equation

$$\frac{dA}{ds} = -\tfrac{1}{2} A \Delta S$$

which can solved explicitly

$$A\big(q(s) \big) = \exp \left[-\frac{1}{2} \int_{s_0}^{s} \Delta S\big(q(s') \big) A(q_0) \, ds' \right].$$

Note that if $A(q) = 0$, so there is zero amplitude to leading order, then $A = 0$ along the entire characteristic emanating from q. Therefore, in the first approximation, the solutions to the wave equation are concentrated on the characteristics; this reflects the fact that waves and signals propagate along characteristics.

If the index of refraction $n(t, \mathbf{x})$ depends on time, then we can no longer replace the wave equation by its stationary counterpart — the Helmholtz equation. Nevertheless,

we can still investigate the high frequency approximation, assuming that the index of refraction is a slowly varying function of both position and time. We make the ansatz

$$u(\mathbf{x}) = A(t, \mathbf{x}, \omega)\, e^{\,\mathrm{i}\, S(t, \mathbf{x}, \omega)}, \tag{3.7}$$

where the amplitude A and phase S are real. Substituting into the wave equation,

$$\Delta u - \frac{n^2}{c^2} u_{tt} = \left[\omega^2 \left(-|\nabla S|^2 + \frac{n^2}{c^2} S_t^2 \right) + \cdots \right] A\, e^{\,\mathrm{i}\, \omega S},$$

where the omitted terms are lower order in ω. Thus, we derive, as a first approximation, the time-dependent form of the *eikonal equation*

$$|\nabla S|^2 = \frac{n^2}{c^2} S_t^2. \tag{3.8}$$

In particular, if n is independent of t, then we can reduce to the time-independent form (3.5) by setting $S(t, \mathbf{x}) = S(\mathbf{x}) - c\,t$.

Maxwell's Equations

A similar derivation holds for *Maxwell's equations* for electromagnetic waves. The electromagnetic field is prescribed by the *electric vector* $\mathbf{E}$ and the *magnetic vector* $\mathbf{H}$. In flat space, these satisfy the system of first order partial differential equations

$$\begin{aligned}
\frac{\partial}{\partial t}(\varepsilon\, \mathbf{E}) - c\,\operatorname{curl} H &= 4\pi\, \mathbf{j}, & \operatorname{div}(\varepsilon\, \mathbf{E}) &= 4\pi\, \rho, \\[2mm]
\frac{\partial}{\partial t}(\mu\, \mathbf{H}) + c\,\operatorname{curl} \mathbf{E} &= 0, & \operatorname{div}(\mu\, \mathbf{H}) &= 0,
\end{aligned} \tag{3.9}$$

where c is the *speed of light*, ε is the *electric permittivity* or *dielectric constant*, μ the *magnetic permittivity* or *permeability*, $\mathbf{j}$ is the *electric current density*, and ρ the *electric charge density*. Assuming that there are no charges and currents, and that the permittivities are constant in time, we derive the time-independent form of Maxwell's equations

$$\operatorname{curl} \mathbf{H} + \mathrm{i}\,\omega\, \mathbf{E} = 0, \qquad \operatorname{curl} \mathbf{E} - \mathrm{i}\,\omega\, \mathbf{H} = 0, \qquad \operatorname{div}(\varepsilon\, \mathbf{E}) = 0, \qquad \operatorname{div}(\mu\, \mathbf{H}) = 0, \quad (3.10)$$

for a simple harmonic field

$$\mathbf{E}(t, \mathbf{x}) = \mathbf{E}(\mathbf{x})\, e^{-\,\mathrm{i}\, c\,\omega\, t}, \qquad \mathbf{H}(t, \mathbf{x}) = \mathbf{H}(\mathbf{x})\, e^{-\,\mathrm{i}\, c\,\omega\, t}.$$

In the case of a homogeneous medium, so that the permittivities ε, μ, are constant, the solutions are superpositions of plane waves

$$\mathbf{E}(t, \mathbf{x}) = \mathbf{e}_0\, e^{\,\mathrm{i}\,\omega(\mathbf{k}\cdot\mathbf{x} - c\,\omega\, t)}, \qquad \mathbf{H}(t, \mathbf{x}) = \mathbf{h}_0\, e^{\,\mathrm{i}\,\omega(\mathbf{k}\cdot\mathbf{x} - c\,\omega\, t)},$$

where $\mathbf{e}_0, \mathbf{h}_0$ are constant vectors, satisfying

$$\varepsilon\,\omega\, \mathbf{e}_0 + \mathbf{h}_0 \wedge \mathbf{k} = 0, \qquad \mu\,\omega\, \mathbf{h}_0 - \mathbf{e}_0 \wedge \mathbf{k} = 0, \qquad \mathbf{e}_0 \cdot \mathbf{k} = \mathbf{h}_0 \cdot \mathbf{k} = 0,$$

where $\wedge$ denotes the *vector cross product*. The vectors $\mathbf{e}_0, \mathbf{h}_0$ which determine the direction of the electric and magnetic field are orthogonal to the direction of propagation $\mathbf{k}$, and to each other. Substituting the second equation into the first, we must have

$$\varepsilon\,\mu\,\omega^2\,\mathbf{e}_0 + (\mathbf{e}_0 \wedge \mathbf{k}) \wedge \mathbf{k} = 0.$$

Now $\mathbf{e}_0$ is orthogonal to $\mathbf{k}$, so

$$(\mathbf{e}_0 \wedge \mathbf{k}) \wedge \mathbf{k} = -\,|\,\mathbf{k}\,|^2\,\mathbf{e}_0.$$

Therefore, for the four equations to be compatible, the wave number $\mathbf{k}$ and the frequency ω must be related by the dispersion relation

$$|\,\mathbf{k}\,| = n\,\omega, \qquad \text{where} \qquad n = \sqrt{\varepsilon\,e\,\mu} \tag{3.11}$$

is the *refractive index* of the medium. Note that the speed of the electromagnetic wave is $v = c\,\omega/|\,\mathbf{k}\,|$, hence $n = c/v$ as it should be.

As in the case of the scalar wave equation, we consider the case when the frequency is large in comparison to the variation in the refractive index n, and so look for solutions of the form

$$\mathbf{E}(\mathbf{x}) = \mathbf{e}(\mathbf{x},\omega)\,e^{\,\mathrm{i}\,\omega S(\mathbf{x},\omega)}, \qquad \mathbf{H}(\mathbf{x}) = \mathbf{h}(\mathbf{x},\omega)\,e^{\,\mathrm{i}\,\omega S(\mathbf{x},\omega)}, \tag{3.12}$$

where the amplitudes $\mathbf{e}, \mathbf{h}$ may be complex, but the phase S is real. Note that

$$\operatorname{curl}\mathbf{E} = (\operatorname{curl}\mathbf{e} + \mathrm{i}\,\omega\,\nabla S \wedge \mathbf{e})\,e^{\,\mathrm{i}\,\omega S}, \qquad \operatorname{div}\mathbf{E} = (\operatorname{div}\mathbf{e} + \mathrm{i}\,\omega\,\nabla S \cdot \mathbf{e})\,e^{\,\mathrm{i}\,\omega S}.$$

Substituting (3.12) into Maxwell's equations, we find the equivalent system

$$\varepsilon\,\mathbf{e} + \nabla S \wedge \mathbf{h} = -\,\frac{1}{\mathrm{i}\,\omega}\operatorname{curl}\mathbf{h}, \qquad\qquad \mu\,\mathbf{h} - \nabla S \wedge \mathbf{e} = \frac{1}{\mathrm{i}\,\omega}\operatorname{curl}\mathbf{e},$$

$$\mathbf{e}\cdot\nabla S = -\,\frac{1}{\mathrm{i}\,\omega}\,(\mathbf{e}\cdot\nabla\log\varepsilon + \operatorname{div}\mathbf{e}), \qquad \mathbf{h}\cdot\nabla S = -\,\frac{1}{\mathrm{i}\,\omega}\,(\mathbf{h}\cdot\nabla\log\mu + \operatorname{div}\mathbf{h}).$$

In the high frequency approximation, we make an asymptotic expansion of the amplitude

$$\mathbf{e}(\mathbf{x},\omega) = \mathbf{e}(\mathbf{x}) + \omega^{-1}\mathbf{e}_1(\mathbf{x}) + \omega^{-2}\mathbf{e}_2(\mathbf{x}) + \cdots$$

and the phase

$$S(\mathbf{x},\omega) = S(\mathbf{x}) + \omega^{-1}S_1(\mathbf{x}) + \omega^{-2}S_2(\mathbf{x}) + \cdots$$

in decreasing powers of the frequency ω. The highest order terms in the system are

$$\varepsilon\,\mathbf{e} + \nabla S \wedge \mathbf{h} = 0, \qquad \mu\,\mathbf{h} - \nabla S \wedge \mathbf{e} = 0, \qquad \mathbf{e}\cdot\nabla S = 0, \qquad \mathbf{h}\cdot\nabla S = 0.$$

If we substitute the second equation into the first, and use the third equation, we find that

$$\varepsilon\,\mu\,\mathbf{e} = -\,\nabla S \wedge (\nabla S \wedge \mathbf{e}) = |\,\nabla S\,|^2\,\mathbf{e}.$$

Therefore we deduce that S must satisfy the same eikonal equation

$$|\,\nabla S\,|^2 = n^2.$$

An alternative approach is to use the second order forms of Maxwell's equations

$$\mathbf{E}_{tt} - \frac{c^2}{n^2}\Delta\mathbf{E} = \frac{c^2}{n^2}\big[\,\nabla(\mathbf{E}\cdot\nabla\log\varepsilon) + \nabla\log\mu \wedge \operatorname{curl}\mathbf{E}\,\big],$$

$$\mathbf{H}_{tt} - \frac{c^2}{n^2}\Delta\mathbf{H} = \frac{c^2}{n^2}\big[\,\nabla(\mathbf{H}\cdot\nabla\log\mu) + \nabla\log\varepsilon \wedge \operatorname{curl}\mathbf{H}\,\big].$$

(3.13)

Working as above also leads to the eikonal equation. The corresponding transport equations can be derived slightly more directly here; they are

$$\frac{d\mathbf{e}}{ds} = \frac{1}{2}\left(\frac{d\log m}{ds} - \Delta S\right)\mathbf{e} - (\nabla\log n\cdot\mathbf{e})\nabla S,$$

$$\frac{d\mathbf{h}}{ds} = \frac{1}{2}\left(\frac{d\log\varepsilon}{ds} - \Delta S\right)\mathbf{h} - (\nabla\log n\cdot\mathbf{h})\nabla S,$$

(3.14)

where d/ds denotes derivatives along the characteristics to the eikonal equation, i.e.,

$$\frac{d\Phi}{ds} = \nabla S\cdot\nabla\Phi.$$

As before, the waves and signals propagate along characteristics.

High Frequency Limit and Quantization

In general, suppose we have a linear partial differential equation

$$\mathcal{F}[\psi] = 0,$$

(3.15)

depending on a large parameter ω. The differential operator $\mathcal{F}$ can be written as

$$\mathcal{F} = F(x, \mathrm{i}\,\partial, \omega)$$

where $F(x, p, \omega)$ is a smooth function, which is a polynomial in the derivative coordinate $p = \mathrm{i}\,\partial$. We use $\partial = (\partial_1, \dots, \partial_n)$ to denote the derivative operators $\partial_j = \partial/\partial x_j$, $j = 1, \dots, n$. There is a problematic ambiguity in this representation, since we have to specify the order of the derivatives and the function. For instance, if $F(x, p) = xp$, then there is a question as to whether this should represent the differential operator $\mathrm{i}\,x\partial$ or $\mathrm{i}\,\partial\cdot x = \mathrm{i}\,x\partial + \mathrm{i}$. For convenience, we adopt the convention that the derivatives always appear last, so xp corresponds to the differential operator $\mathrm{i}\,x\partial$. (However, this will come back to haunt us later.)

More generally, if F is no longer a polynomial in the derivative variable p, then we can regard[†]

$$\mathcal{F}[\psi] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(x, p, \omega)\,\widehat{\psi}(p)\,e^{\mathrm{i}\,px}\,dp.$$

(3.16)

[†] Here, for simplicity, we specialize to the one-dimensional case, with the evident multidimensional generalization.

as a *Fourier integral operator*. Here

$$\widehat{\psi}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(p) \, e^{-\mathrm{i}\,p\,x} \, dx \qquad (3.17)$$

denotes the Fourier transform of $\psi(x)$, assuming $\psi \in \mathrm{L}^2$. The convergence of the integral (3.16) requires that the integrand F satisfy certain global growth conditions.

In the high frequency limit, we make the ansatz

$$\psi(x,\omega) = A(x,\omega) \, e^{\mathrm{i}\,\omega S(x,\omega)},$$

where A and S are real, A having the usual asymptotic expansion in decreasing powers of ω. In order to determine the analogue of the eikonal equation, it is helpful to rewrite the operator in the form

$$\mathcal{F} = F\left(x, \frac{1}{\mathrm{i}\,\omega}\,\partial, \omega\right),$$

(where F has changed meanings). We assume that for large ω we can expand F in an asymptotic series

$$F(x,p,\omega) \;\sim\; F_n(x,p)\,\omega^n + F_{n-1}(x,p)\,\omega^{n-1} + \,\cdots\,,$$

where we call F_n the leading component of the operator. Now, note that if ψ is given as above, then

$$\frac{1}{\mathrm{i}\,\omega}\,\frac{\partial\psi}{\partial x_j} = \frac{1}{\mathrm{i}\,\omega}\,\frac{\partial}{\partial x_j}\left(A\,e^{\mathrm{i}\,\omega S}\right) = \frac{\partial S}{\partial x_j}\,\psi + \frac{1}{\mathrm{i}\,\omega}\,\frac{\partial \log A}{\partial x_j}\,\psi.$$

But the second term has order less that the first. It is not difficult to see that, in general,

$$\mathcal{F}[\psi] = F\left(x, \frac{1}{\mathrm{i}\,\omega}\,\partial, \omega\right)\psi = \omega^n F_n(x, \nabla S)\,\psi + \mathrm{O}(\omega^{n-1}).$$

Therefore, in the high frequency limit, the term $\omega^n F_n(x, \nabla S)\psi$ will dominate the asymptotic expansion. In order that the equation $\mathcal{F}[\psi] = 0$ hold, then, we find the analogue of the eikonal equation to be

$$F_n(x, \nabla S) = 0. \qquad (3.18)$$

For instance, the Helmholtz equation is already homogeneous of degree 0, corresponding to $F(x,p,\omega) = -|p|^2 + n^2$. Thus (3.18) is coincides with the eikonal equation (3.5).

Now, what about the Hamilton–Jacobi equation? It will be the high frequency limit of some linear wave equation. In fact, if we set

$$F(t, x, \pi, p) = \pi + H(t, x, p),$$

where H is the Hamiltonian function, then

$$F(t, x, S_t, \nabla S) = S_t + H(t, x, \nabla S) = 0$$

is the Hamilton–Jacobi equation. This indicates that the corresponding wave equation is

$$\frac{1}{\mathrm{i}\,\omega}\,\psi_t + \mathcal{H}[\psi] = 0, \qquad \text{where} \qquad \mathcal{H} = H\left(t, x, \frac{1}{\mathrm{i}\,\omega}\,\partial\right).$$

We denote $\omega^{-1} = \hbar$, which is known as *Planck's constant* and has the units of inverse frequency. (This is motivated by Einstein's formula $E = \hbar\omega$, where we fix the energy E.) The resulting equation

$$i\,\hbar\,\psi_t = \mathcal{H}[\psi].$$

is known as the *Schrödinger equation*, and is *the* fundamental equation of quantum mechanics. We have found it by requiring that its high frequency (low $\hbar$) limit reduce to the Hamilton–Jacobi equation of classical mechanics. This means that we are endowing classical particles with a wave-like interpretation. The differential operator $\mathcal{H} = H(t, x, -i\,\hbar\,\partial)$ is known as the *Hamiltonian operator* for the quantum mechanical Schrödinger equation.

Example 3.1. Consider the case of a particle in a central force field. Here the Hamiltonian is given by

$$H(p, q) = \frac{|p|^2}{2m} + V(r), \tag{3.19}$$

where $r = |q|$. (Recall that this also describes the motion of two interacting masses, provided we go to center of mass coordinates.) The corresponding Hamilton–Jacobi equation is

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left|\frac{\partial S}{\partial q}\right|^2 + V(r) = 0. \tag{3.20}$$

Replacing p_j by $-i\,\partial_j$ produces the associated Hamiltonian operator

$$\mathcal{H} = -\frac{\hbar^2}{2m}\,\Delta + V(r), \tag{3.21}$$

where Δ is the ordinary Laplacian. The Schrödinger equation is

$$i\,\hbar\,\psi_t = -\frac{\hbar^2}{2m}\,\Delta\psi + V(r)\,\psi. \tag{3.22}$$

In the case $V(r) = -e^2/r$, where e is the charge on an electron, we are in the situation of the quantum mechanical *hydrogen atom*, meaning a single electron circling a single (heavy) proton.

A Word of Warning: The Schrödinger equation looks a lot like the heat equation, but the complex factor i makes it of an entirely different character. It is, in fact, a dispersive hyperbolic partial differential equation, not a dissipative parabolic equation. One way to see the difference right away is to look at the norm of the solution[†] $|\psi|^2 = \psi\,\overline{\psi}$. For the one-dimensional Schrödinger equation[‡]

$$i\,\hbar\,\psi_t = -\frac{\hbar^2}{2m}\,\psi_{xx} + V(x)\,\psi, \tag{3.23}$$

[†] We use an overbar to denote complex conjugate throughout.

[‡] The proof can be easily generalized to the n-dimensional case; see below.

we have

$$\frac{\partial}{\partial t} \, |\,\psi\,|^2 = \psi\,\overline{\psi}_t + \psi_t\,\overline{\psi} = \psi\,\left(\frac{\hbar}{2\,\mathrm{i}\,m}\,\overline{\psi}_{xx} - V(x)\,\overline{\psi}\right) + \left(-\frac{\hbar}{2\,\mathrm{i}\,m}\,\psi_{xx} + V(x)\,\psi\right)\overline{\psi}$$

$$= \frac{\hbar}{2\,\mathrm{i}\,m}\,\frac{\partial}{\partial x}\left(\psi\,\overline{\psi}_x - \overline{\psi}\,\psi_x\right).$$

Therefore, the L^2 norm of $\psi(x)$, such that it and its first derivative tend to 0 as $x \to \infty$, is constant:

$$\frac{d}{dt}\,\|\,\psi\,\|^2 = \frac{d}{dt}\int_{-\infty}^{\infty} |\,\psi\,|^2\,dx = \frac{\hbar}{2\,\mathrm{i}\,m}\left(\psi\,\overline{\psi}_x - \psi_x\,\overline{\psi}\right)\bigg|_{-\infty}^{\infty} = 0.$$

This is in contrast with the heat equation, where the L^2 norm of solutions decreases as $t^{-1/2}$ and they are immediately smoothed out.

At the moment, things seem rather simple. However, there are genuine problems with the passage from a classical mechanical system to its corresponding quantum mechanical counterpart. As an example, consider the central force field problem in $\mathbb{R}^2$, but written in polar coordinates. The classical Hamiltonian becomes

$$\mathcal{H} = \frac{1}{2\,m}\left(p_r^2 + \frac{1}{r^2}\,p_\theta^2\right) + V(r),$$

where p_r, p_θ are the momenta conjugate to the r, θ coordinates. The Hamilton–Jacobi equation is

$$\frac{\partial S}{\partial t} + \frac{1}{2\,m}\left[\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial S}{\partial \theta}\right)^2\right] + V(r) = 0.$$

If we write the corresponding Schrödinger equation, we have

$$\mathrm{i}\,\hbar\,\psi_t = -\frac{\hbar^2}{2\,m}\left(\psi_{rr} + \frac{1}{r^2}\,\psi_{\theta\theta}\right) + V(r)\,\psi. \tag{3.24}$$

On the other hand, the rectangular coordinate Schrödinger equation involves the Laplacian, which, in polar coordinates, assumes a slightly different form:

$$\Delta\psi = \psi_{xx} + \psi_{yy} = \psi_{rr} + \frac{1}{r}\,\psi_r + \frac{1}{r^2}\,\psi_{\theta\theta}, \tag{3.25}$$

containing a first order term not present in (3.24). Thus, although the two classical systems are completely equivalent under the change of variables, the same is not true for the two corresponding Schrödinger equations.

What is going on? In our derivation of the geometric optics approximation to wave optics, we only looked at the leading order terms in the partial differential equation, and ignored lower order terms in $\omega = \hbar^{-1}$. Therefore, many different equations will reduce down to the same classical system in the high frequency limit, and we have no way of knowing from classical mechanics alone what the correct lower order terms are. (These are, as I understand it, even difficult to determine experimentally in quantum mechanical systems.) Thus, there is an inherent ambiguity in our derivation of the Schrödinger equation. The

point is that changes of coordinates will preserve the leading order terms, but will not preserve the lower order components (even if there are none in one particular coordinate system). Thus, the rectangular Schrödinger equation, re-expressed in polar coordinates has the form

$$0 = \left[-i\hbar\,\frac{\partial}{\partial t} + \frac{1}{2m}\left(-i\hbar\,\frac{\partial}{\partial r} \right)^2 + \frac{1}{2mr^2}\left(-i\hbar\,\frac{\partial}{\partial \theta} \right)^2 + V(r) \right]\psi$$
$$+ \frac{\hbar}{2mr}\left(-i\hbar\,\frac{\partial}{\partial r} \right)\psi.$$

The leading order terms agree with the polar coordinate Schrödinger equation (3.24), but the final order $\hbar$ term, which does not appear in the classical limit, is absent.

For the early quantum mechanists, this was not viewed as a problem. Basically, they required one to quantize only in rectangular coordinates only. However, this is far from a relativistic viewpoint, which asserts that physical laws must be independent of any particular coordinates system on the space-time manifold. And, to this day, this inherent ambiguity in quantization is still causing problems with the mathematical foundations of quantum mechanics.

This is also related to our earlier ambiguity about the ordering of functions and derivatives when we replaced the Hamiltonian function by the corresponding Hamiltonian operator. In fact, note that the commutator

$$[q, -i\hbar\,\partial_q] = q\,(-i\hbar\,\partial_q) - (-i\hbar\,\partial_q)\,q = i\hbar$$

of the operators corresponding to q and p is of higher order in $\hbar$ than the operators themselves. More generally, we have can show the following.

Proposition 3.2. *Let* $F(q,p), H(q,p)$ *be smooth functions of* q *and polynomial in* p. *Let* $\mathcal{F}, \mathcal{H}$ *be the corresponding differential operators. Then*

$$[\mathcal{F}, \mathcal{H}] = i\hbar\,\mathcal{G} + \mathrm{O}(\hbar^2),$$

where $\mathcal{G}$ *is the differential operator obtained from their Poisson bracket* $G = \{F, H\}$.

Proof: Here we present the proof in one dimension; the multi-dimensional version is analogous. By linearity, it suffices to check the case when

$$F(q,p) = f(q)\,p^m, \qquad H(q,p) = h(q)\,p^n,$$

are monomials. Then

$$G = \{F, H\} = F_q\,H_p - F_p\,H_q = \left[\,n\,f'(q)\,h(q) - m\,f(q)\,h'(q)\,\right] p^{n+m-1}.$$

On the other hand, the corresponding operators are

$$\mathcal{F} = f(q)\,(-i\hbar\,\partial_q)^m, \qquad \mathcal{H} = h(q)\,(-i\hbar\,\partial_q)^n,$$
$$\mathcal{G} = \left[\,n\,f'(q)\,h(q) - m\,f(q)\,h'(q)\,\right](-i\hbar\,\partial_q)^{n+m-1}.$$

On the other hand, the commutator operator is

$$[\mathcal{F}, \mathcal{H}] = \left[\,m\,f(q)\,h'(q) - n\,f'(q)\,h(q)\,\right](-i\hbar)^{n+m}\partial_q^{n+m-1} + \cdots,$$

where the omitted terms have the form $g_j(q)(-i\hbar)^{n+m}\partial_q^{n+m-j}$ for $j \geq 2$. Comparison of the two expressions establishes the result.   Q.E.D.

A similar result holds for the Fourier integral operators defined by non-polynomial Hamiltonians. Therefore, if we quantize the function $F(q,p)$ by using a different ordering of the positions and momenta, we obtain a different Hamiltonian operator, but these will all differ by commutators. For instance, if $H(q,p) = h(q)p^n$, then we can rewrite this as $f(q)p^j g(q) p^{n-j}$, say, where $fg = h$, leading to the alternative Hamiltonian operator

$$\widetilde{\mathcal{H}} = f(q)(-i\hbar\partial_q)^j g(q)(-i\hbar\partial_q)^{n-j}.$$

But

$$\mathcal{H} - \widetilde{\mathcal{H}} = f(q)\left[(-i\hbar\partial_q)^j, g(q)\right](-i\hbar\partial_q)^{n-j} = O(\hbar)$$

by Proposition 3.2. A similar result holds for the Hamiltonian operators obtained by re-expressing the Hamiltonian function in different coordinate systems. This is because if we change $\overline{q} = \varphi(q)$, then $\overline{p} = \varphi'(q)^{-1}p$, and $\overline{\partial}_q = \varphi'(q)^{-1}\partial_q$. However, if $H(q,p) = h(q)p^n$, then

$$\overline{H} = h(\varphi^{-1}(\overline{q}))\,\varphi'(\varphi^{-1}(\overline{q}))^{-n}\,\overline{p}^n,$$

leading to

$$\overline{\mathcal{H}} = h(\varphi^{-1}(\overline{q}))\,\varphi'(\varphi^{-1}(\overline{q}))^{-n}\,(-i\hbar\overline{\partial}_q)^n.$$

whereas transforming the original Hamiltonian operator gives

$$h(\varphi^{-1}(\overline{q}))\left(-i\hbar\,\varphi'(\varphi^{-1}(\overline{q}))^{-1}\,\partial_q\right)^n.$$

But these two operators again differ by commutators, and so agree to leading order.

Corollary 3.3. *The quantum operators obtained by reordering the q and $-i\hbar\partial_q$ factors in the formula for $H(q,p)$, or by changing variables, differ by terms involving commutators that are of order $\hbar$ and higher.*

A significant problem, which is still not fully resolved, is whether there is a consistent choice of quantum Hamiltonian operator. Dirac decided that the consistency might be a consequence of requiring that the commutators and Poisson brackets match up exactly: If $F(q,p), H(q,p)$, are smooth functions of q, polynomial in p with the Poisson bracket $\{F, H\} = G$, then we require that the corresponding quantum Hamiltonian operators satisfy the commutator relationship

$$[\mathcal{F}, \mathcal{H}] = i\hbar\mathcal{G}.$$

There are some mathematically reasonable ways to do this, most notably the method of geometric quantization, [**59**]. However, some elementary systems, e.g., the helium atom, are still not covered by this approach, and it has not been significantly developed in recent years. Moreover, if we impose a few additional reasonable assumptions, then we run into a roadblock: there are no consistent methods of quantization satisfying this hypothesis. Another approach is to allow terms of order $O(\hbar^2)$ in the commutator. Again, there are some partial results, but the full theory still seems rather far off.

Now, return to the Schrödinger equation

$$i\,\hbar\,\psi_t = \mathcal{H}[\psi]. \tag{3.26}$$

Assume that the Hamiltonian operator $\mathcal{H}$ is independent of t. Then we can separate variables by setting

$$\psi(t,x) = \widehat{\psi}(x)\,e^{\,i\,\omega t},$$

leading to the time independent form of the Schrödinger equation

$$\mathcal{H}[\widehat{\psi}\,] = \hbar\,\omega\,\widehat{\psi}.$$

If we use Einstein's relation between frequency and energy

$$E = \hbar\,\omega, \tag{3.27}$$

we see that the energy of a system must be an *eigenvalue* of the corresponding Hamiltonian operator. There is thus an intimate connection between the possible energies of a physical system, and the *spectrum* of the corresponding Hamiltonian operator. We assume (for the time being) that the solutions of the time-independent Schrödinger equation must be smooth and bounded over all space. In the particular case of the hydrogen atom, or a more general particle in a central force field $V(r)$, with $V \to 0$ as $r \to \infty$, the spectrum of

$$\mathcal{H} = -\Delta + V(r)$$

consists of two parts:

(i) The *discrete spectrum*, $E < 0$, which consists of a finite number of negative eigenvalues corresponding to *bound states*. The associated eigenfunctions ψ are in L^2, and, in particular, $\psi \to 0$ as $r \to \infty$.

(ii) The *continuous spectrum*, $E > 0$, where the associated eigenfunction ψ no longer goes to zero as $r \to \infty$, but rather its asymptotic behavior is like that of a plane wave $e^{\,i\,\mathbf{k}\cdot\mathbf{x}}$. These correspond to *scattering states*.

The key difference between classical mechanics and quantum mechanics is that in classical mechanics, the energy can take on any positive value, but in quantum mechanics, the bound state energies are quantized, i.e. they can only take on discrete values. The investigation of the spectrum of Hamiltonian operators is the fundamental problem of quantum mechanics. To do this properly, we need to discuss the basics of quantum mechanics and the theory of operators on Hilbert space.

4. Basics of Quantum Mechanics.

	Classical mechanics	**Quantum mechanics**
Underlying space	Phase space M	Complex Hilbert space $\mathcal{E}$
State of system	Point $(p, q) \in M$	Wave function $0 \neq \psi \in \mathcal{E}$
Observable	Function F on M	Self-adjoint operator $\mathcal{F}$ on $\mathcal{E}$
Result of Measurement	Deterministic; always $F(p, q)$	Probabilistic; on average $\langle\,\psi\,,\mathcal{F}\psi\,\rangle$
Possible values of physical quantity	Any real number	Spectrum of $\mathcal{F}$
Time evolution (dynamics)	Hamilton's equations	Schrödinger equation
Change in observable	Poisson bracket $\{\,F, H\,\}$	Commutator $[\,\mathcal{F}, \mathcal{H}\,]$
Change of coordinates	Canonical transformation	Unitary transformation

Phase space

Classical mechanics takes place in phase space $M = \mathbb{R}^n \times \mathbb{R}^n$, with coordinates $(p, q) = (p_1, \ldots, p_n, q_1, \ldots, q_n)$. In Schrödinger's wave mechanics, the corresponding quantum mechanical phase space will be the complex Hilbert space $\mathcal{E} = \mathrm{L}^2(\mathbb{R}^n)$, whose elements are complex-valued square-integrable functions $\psi \colon \mathbb{R}^n \to \mathbb{C}$ depending on the position variables q and satisfying

$$\|\,\psi\,\|^2 = \langle\,\psi\,,\psi\,\rangle = \int_{\mathbb{R}^n} \psi\,\overline{\psi}\,dq = \int_{\mathbb{R}^n} |\,\psi\,|^2\,dq < \infty. \tag{4.1}$$

A nonzero element $\psi \neq 0$ is called a *wave function* of the quantum mechanical system, and completely determines the dynamical state of the system at a given time. In quantum mechanics, ψ is interpreted probabilistically; roughly speaking, it measures the (relative) probability of finding the system in the particular physical state q. For instance, if we have just one particle, then this will give the probability that the particle is in the position q of configuration space. More specifically, if we normalize ψ to have unit norm $\|\,\psi\,\| = 1$, then the probability density $|\,\psi(q)\,|^2\,dq$ represents the probability that the system will belong to the element dq of configuration space. Note that a normalized wave function is determined to within an arbitrary phase factor $e^{\mathrm{i}\theta}$ for $\theta \in \mathbb{R}$.

More generally, the phase space of a quantum mechanical system is some (abstract) *complex Hilbert space* $\mathcal{E}$, which, by definition, is a complete inner product space, the simplest example being $\mathcal{E} = \mathrm{L}^2(\mathbb{R}^n)$.

Observables

The Hamiltonian operator for a system is an example of an observable physical quantity for the system, representing its *energy*. In general, an *observable* physical quantity is given by a linear operator

$$L : \mathcal{E} \longrightarrow \mathcal{E}.$$

defined on the Hilbert space. Actually, almost all operators of interest are not defined everywhere on $\mathcal{E}$, but only on some dense subspace of $\mathcal{E}$.

For instance, let us focus on the one-dimensional case, where $\mathcal{E} = \mathrm{L}^2(\mathbb{R})$. The *position operator* Q corresponding to the classical observable q is given by multiplication by q:

$$Q[\psi(q)] = q \, \psi(q). \tag{4.2}$$

However, if $\psi(q)$ is an L^2 function, it is not necessarily true that $q \, \psi(q)$ is in L^2. Thus the domain $\mathcal{D}(Q)$ of the position operator consists of all functions $\psi \in \mathrm{L}^2$ such that $q \, \psi(q) \in \mathrm{L}^2$ also. Clearly this is a dense subspace of L^2. (Indeed, it includes all smooth functions of compact support, which already form a dense subspace of L^2.) Similarly, the *momentum operator* P corresponding to the classical observable p is given by differentiation:

$$P[\psi(q)] = - \, \mathrm{i} \, \hbar \, \frac{\partial \psi}{\partial q}. \tag{4.3}$$

It is only defined on the domain

$$\mathcal{D}(P) = \left\{ \, \psi \in \mathrm{L}^2 \ \middle| \ \psi \in \mathrm{C}^1, \ \frac{\partial \psi}{\partial q} \in \mathrm{L}^2 \, \right\},$$

which is also a dense subspace. One can also extend P to the *Sobolev space* $\mathrm{H}^1 \subset \mathrm{L}^2$ consisting of functions $\psi \in \mathrm{L}^2$ whose (weak) derivative $\partial \psi / \partial q \in \mathrm{L}^2$ also, cf. [**3**].

If the system is in a state ψ contained in the domain of the observable L, the *expected value* of the observable is, by definition, its average:

$$\langle \, L \, \rangle = \frac{\langle \, \psi \, , L \, \psi \, \rangle}{\| \, \psi \, \|^2}. \tag{4.4}$$

This formula is motivated by considering specific examples, e.g., the position and momentum operators in the Schrödinger representation. If we assume that an observable physical quantity can only assume real values, then its mean value must also be real. Therefore,

$$\langle \, \psi \, , L \, \psi \, \rangle = \langle \, L \, \rangle = \overline{\langle \, L \, \rangle} = \overline{\langle \, \psi \, , L \, \psi \, \rangle} = \langle \, L \, \psi \, , \psi \, \rangle = \langle \, \psi \, , L^* \psi \, \rangle.$$

Here L^* denotes the Hermitian adjoint of L, which, for simplicity, we assume to be defined on the same domain as L. Since this holds for a dense set of states $\psi \in \mathcal{E}$, we conclude that L must be a *self-adjoint operator*, meaning that $L^* = L$, and hence

$$\langle \, \psi \, , L\chi \, \rangle = \langle \, L \, \psi \, , \chi \, \rangle = \overline{\langle \, \chi \, , L \, \psi \, \rangle}.$$

To prove this, we let $\alpha \in \mathbb{C}$, and so

$$\langle \, \psi + \alpha \chi \, , L(\psi + \alpha \chi) \, \rangle = \langle \, \psi \, , L \, \psi \, \rangle + \overline{\alpha} \, \langle \, \chi \, , L \, \psi \, \rangle + \alpha \, \langle \, \psi \, , L \chi \, \rangle + | \, \alpha \, |^2 \, \langle \, \chi \, , L \chi \, \rangle$$

must be real. The first and last summands are real by assumption, so the only way that the entire expression can be real for all α is if the self-adjoint property holds.

Thus, observable physical quantities are given by self-adjoint linear operators on the underlying Hilbert space. For example, in the case of wave mechanics, $\mathcal{E} = \mathrm{L}^2(\mathbb{R}^n)$. It is easy to check that, for all j, k, the *position operators*

$$Q_j\big[\,\psi(q)\,\big] = q_j\,\psi(q), \tag{4.5}$$

the *momentum operators*

$$P_j\big[\,\psi(q)\,\big] = -\,\mathrm{i}\,\hbar\,\frac{\partial\psi}{\partial q_j}, \tag{4.6}$$

and the *angular momentum operators*

$$J_{jk}\big[\,\psi(q)\,\big] = -\,\mathrm{i}\,\hbar\left(q_j\,\frac{\partial\psi}{\partial q_k} - q_k\,\frac{\partial\psi}{\partial q_j}\right), \tag{4.7}$$

are "formally self adjoint" operators, i.e., (3.15) holds for ψ in a suitable dense subdomain of $\mathcal{E}$. The fully rigorous definition of self-adjointness will be delayed.

Schrödinger's Equation

By the superposition principle, the time evolution of a quantum mechanical system must be governed by a linear wave equation. If we also assume that knowledge of the initial wave function ψ at $t = 0$ is sufficient to determine the time evolution of the system, then the equation must be first order in the time variable:

$$\psi_t = \mathcal{M}[\psi],$$

where M is some operator. Moreover, assume that the equation preserves the norm; then

$$0 = \frac{d}{dt}\,\|\,\psi\,\|^2 = \langle\,\psi_t\,,\psi\,\rangle + \langle\,\psi\,,\psi_t\,\rangle = \langle\,\mathcal{M}\psi\,,\psi\,\rangle + \langle\,\psi\,,\mathcal{M}\psi\,\rangle = \langle\,\psi\,,(\mathcal{M} + \mathcal{M}^*)\psi\,\rangle.$$

Arguing as above, we conclude that

$$\mathcal{M}^* = -\,\mathcal{M}$$

must be skew-adjoint. It is traditional to write

$$\mathcal{M} = \frac{1}{\mathrm{i}\,\hbar}\,\mathcal{H},$$

where $\hbar$ is Planck's constant, and $\mathcal{H}$ is called the *Hamiltonian* of the system, which, owing to the factor i is self-adjoint:

$$\mathcal{H}^* = \mathcal{H}.$$

Thus, the time evolution is governed by the *Schrödinger equation*

$$\mathrm{i}\,\hbar\,\psi_t = \mathcal{H}[\psi]. \tag{4.8}$$

For the moment, we set $\hbar = 1$ without loss of generality, since this can be effected by suitably rescaling the time t.

Theorem 4.1. *If $\mathcal{H}$ is a self-adjoint operator on a Hilbert space $\mathcal{E}$, then it generates a strongly continuous one-parameter group of unitary transformations*

$$U(t) = \exp(-\mathrm{i}\,t\,\mathcal{H}). \tag{4.9}$$

Conversely, if $U(t)$ is any strongly continuous one-parameter unitary group on $\mathcal{E}$, then there is a self-adjoint operator $\mathcal{H}$ on $\mathcal{E}$, such that (4.9) holds.

Unitarity means the operator preserves the inner product and hence the norm, so that, for each real t,

$$\langle\, U(t)\,\psi\,,U(t)\,\chi\,\rangle = \langle\,\psi\,,\chi\,\rangle, \qquad \text{or, equivalently,} \qquad U(t)^* = U(t)^{-1}. \tag{4.10}$$

One-parameter group means that for each real t and s

$$U(t+s) = U(t)\,U(s). \tag{4.11}$$

Strongly continuous means that

$$U(t)\,\psi \;\longrightarrow\; U(t_0)\,\psi \qquad \text{whenever} \qquad t \longrightarrow t_0 \qquad \text{and} \qquad \psi \in \mathcal{E}. \tag{4.12}$$

We can recover $\mathcal{H}$ from $U(t)$ according to the limiting formula

$$\frac{U(t)\,\psi - \psi}{t} \;\longrightarrow\; \mathcal{H}[\psi] \qquad \text{as} \qquad t \longrightarrow 0 \tag{4.13}$$

for ψ in the domain of $\mathcal{H}$. In fact, since $\mathcal{H}$ is self-adjoint, the left hand side approaches a limit if and only if ψ is in the domain of $\mathcal{H}$. We call $-\mathrm{i}\,\mathcal{H}$ the *infinitesimal generator* of the one-parameter unitary group $U(t)$.

In the *Schrödinger picture*, the observables are fixed, and the states of the system evolve in time according to

$$\psi(t) = U(t)\,\psi(0), \tag{4.14}$$

so that the expected value of the observable L is given by

$$\langle\,L\,\rangle(t) = \langle\,\psi(t)\,,L\,\psi(t)\,\rangle = \langle\,U(t)\,\psi\,,L\,U(t)\,\psi\,\rangle = \langle\,\psi\,,U(t)^*L\,U(t)\,\psi\,\rangle. \tag{4.15}$$

Alternatively, in the *Heisenberg picture*, the states are fixed, while the observables evolve in time according to

$$L(t) = U(t)^*L\,U(t) = U(t)^{-1}L\,U(t), \tag{4.16}$$

satisfying the dynamical equation

$$\mathrm{i}\,\hbar\,L_t = [\,\mathcal{H},L\,]. \tag{4.17}$$

Spectrum

Let $L\colon \mathcal{E} \to \mathcal{E}$ be an observable. The statistical fluctuations of the measurement of L when the system is in a state ψ is determined by the *root-mean-square deviation* ΔL,

defined by[†]

$$(\Delta L)^2 \equiv \langle\, (L - \langle\, L\,\rangle)^2\,\rangle = \langle\, L^2\,\rangle + \langle\, L\,\rangle^2. \tag{4.18}$$

When the deviation vanishes, $\Delta L = 0$, there are no fluctuations, and the quantity L takes on a well-defined value, namely $\langle\, L\,\rangle$, with certainty. For this to hold, according to (4.18),

$$\frac{\langle\, \psi\,, L^2\,\psi\,\rangle}{\|\,\psi\,\|^2} = \frac{\langle\, \psi\,, L\,\psi\,\rangle^2}{\|\,\psi\,\|^4}$$

or, since $\langle\, \psi\,, L^2\,\psi\,\rangle = \langle\, L\,\psi\,, L\,\psi\,\rangle = \|\, L\,\psi\,\|^2$, we find

$$\langle\, \psi\,, L\,\psi\,\rangle = \|\,\psi\,\|\,\|\, L\,\psi\,\|.$$

This means that we have equality in the Cauchy–Schwarz inequality, and the only way for this to happen is if ψ and $L\psi$ are parallel, which implies that ψ is an eigenfunction of L:

$$L\psi = \lambda\,\psi \qquad \text{for some scalar } \lambda \in \mathbb{C}.$$

Theorem 4.2. *An observable physical quantity assumes, with certainty, a well-defined value if and only if the state of the system is represented by an eigenfunction of its associated quantum mechanical operator L. In this case, the value of the quantity is the eigenvalue:*

$$\langle\, L\,\rangle = \lambda.$$

Recall that the eigenvalues of a self-adjoint operator are always real. Moreover, the eigenfunctions corresponding to different eigenvalues are orthogonal. We can further assume, without loss of generality, that the eigenfunctions have been normalized: $\|\,\psi\,\| = 1$.

In general, given a state ψ, an observable L will take on values $\lambda \in \mathbb{R}$ with various probabilities. In other words, the observable will determine a certain probability measure $d\mu(\lambda) = d\mu_{L,\psi}(\lambda)$ on the space $\mathbb{R}$ of possible values for L. This measure must satisfy

$$\int_{-\infty}^{\infty} d\mu(\lambda) = 1. \tag{4.19}$$

The *expected value* of L then is the same as the first moment of the corresponding measure:

$$\langle\, L\,\rangle = \int_{-\infty}^{\infty} \lambda\, d\mu(\lambda) = \langle\, \lambda\,\rangle. \tag{4.20}$$

The *deviation* of L is related to the variance of μ:

$$(\Delta L)^2 = \sigma^2 = \int_{-\infty}^{\infty} \big(\lambda - \langle\, \lambda\,\rangle\big)^2 d\mu(\lambda), \tag{4.21}$$

[†] *Note:* $\langle\, L\,\rangle$ is a scalar, so we should write $L - \langle\, L\,\rangle\, \mathrm{I}$ for the operator appearing in the middle, but the identity operator is commonly omitted from such formulae since it is clear from the context.

which, as in (4.18), is calculated from the second moment

$$\langle\, L^2 \,\rangle = \int_{-\infty}^{\infty} \lambda^2 \, d\mu(\lambda) = \langle\, \lambda^2 \,\rangle.$$

In general, the *moments* of the probability measure $d\mu$ are given by the analogous formula

$$\langle\, L^n \,\rangle = \int_{-\infty}^{\infty} \lambda^n \, d\mu(\lambda) = \langle\, \lambda^n \,\rangle, \qquad n = 0, 1, 2, 3, \dots\,.$$

To reconstruct the probability distribution, the easiest way is to use the *characteristic function*, which is essentially the Fourier transform of the measure:

$$\chi(\xi) = \int_{-\infty}^{\infty} e^{-\,\mathrm{i}\,\xi\,\lambda} \, d\mu(\lambda) = \langle\, e^{-\,\mathrm{i}\,\xi\,\lambda} \,\rangle,$$

from which we can reconstruct μ by Fourier inversion. For instance, if we have a discrete probability measure

$$\mu(\lambda_\nu) = c_\nu > 0, \qquad \nu = 1, \dots, n, \qquad \text{with} \qquad \sum_{\nu=1}^{n} c_\nu = 1,$$

then

$$\chi(\xi) = \sum_{\nu=1}^{n} c_\nu e^{-\,\mathrm{i}\,\xi\,\lambda_\nu}.$$

Thus, to determine the probability distribution associated with an observable L in a dynamical state ψ, we need to calculate the expected value $\langle\, e^{\,\mathrm{i}\,\xi\,L} \,\rangle$ of the solution to the Schrödinger equation for L. For instance, suppose

$$\psi = \sum c_\nu \, \psi_\nu$$

is a superposition of states ψ_ν associated with eigenvalues λ_ν. Since L is self-adjoint, the ψ_ν are orthonormal. Assume that ψ is normalized, so that

$$\|\,\psi\,\|^2 = \sum c_\nu^2 = 1.$$

Then

$$e^{-\,\mathrm{i}\,\xi\,L}\,\psi = \sum c_\nu e^{-\,\mathrm{i}\,\xi\,\lambda_\nu}\,\psi_\nu,$$

hence the characteristic function associated with the corresponding probability distribution will be

$$\langle\, e^{\,\mathrm{i}\,\xi\,L} \,\rangle = \langle\, \psi\,,\, e^{-\,\mathrm{i}\,\xi\,L}\,\psi \,\rangle = \sum c_\nu^2 e^{-\,\mathrm{i}\,\xi\lambda_\nu}.$$

We conclude that the associated probability measure is purely discrete, concentrated on the eigenvalues λ_ν, each of which has probability c_ν^2. Therefore, we have the following fundamental principle: for a state built up entirely of discrete eigenfunctions, the only values that an observable may take on are the associated eigenvalues. In particular, we have deduced the basic superposition principle of quantum mechanics: If measurement of

a system in state ψ_1 leads to result 1, while measurement in state ψ_2 leads to result 2, then every linear combination $c_1 \psi_1 + c_2 \psi_2$ gives a state in which the same measurement leads either to result 1 or result 2.

Only very special operators will have a purely discrete spectrum. (An example is the harmonic oscillator, considered below.) Thus, in general, we must widen our scope beyond just the eigenvalues of an operator. For instance, consider the one-dimensional momentum operator $P = -i\hbar\,\partial_q$. If $\psi(q)$ is to be an eigenfunction for P it must satisfy

$$-i\hbar\,\psi_q = \lambda\psi,$$

which implies that ψ is a multiple of the complex exponential $\exp(-i\hbar^{-1}\lambda q)$. But this function, while bounded, is not an L^2 function, and thus does not lie in our Hilbert space. Things are even stranger for the position operator $Q = q$. If $\psi(q)$ is to be an eigenfunction for Q, it must satisfy

$$q\,\psi(q) = \lambda\,\psi(q),$$

for some constant λ. But this implies that $\psi(q) = 0$ except possibly at $q = 0$. This would imply that $\psi = 0$ as an L^2 function, which is not normalizable. Therefore, we would conclude that the position operator cannot have physical values, which is clearly untenable.

There are two ways out of this dilemma. A physicist would, at this point, introduce the *Dirac delta function*, which is a "function" $\delta(q)$ which satisfies

$$\delta(q) = 0, \qquad q \neq 0, \qquad \text{whereas} \qquad \int_{-\infty}^{\infty} \delta(q)\,dq = 1. \tag{4.22}$$

In point of fact, there is no classical functions that simultaneously satisfies both of these properties. More rigorously, $\delta(q)$ defines a *distribution* or *generalized function*, namely a continuous linear functional on the space C_0^∞ of smooth functions with compact support, satisfying

$$\langle\,\delta\,,f\,\rangle = \int_{-\infty}^{\infty} f(q)\,\delta(q)\,dq = f(0), \qquad \text{for all} \qquad f \in C_0^\infty. \tag{4.23}$$

An alternative approach is to note that while the position operator Q does not have any eigenfunctions per se, it does have "approximate eigenfunctions". Indeed, given $\lambda \in \mathbb{R}$, if we set

$$\widehat{\psi}_\sigma(q) = \frac{1}{\pi^{1/4}\,\sigma^{1/2}}\,e^{-(q-\lambda)^2/(2\,\sigma^2)}, \qquad \sigma > 0,$$

so that $\|\,\widehat{\psi}_\sigma\,\| = 1$ for all σ, then

$$\|\,Q\,\widehat{\psi}_\sigma - \lambda\,\widehat{\psi}_\sigma\,\|^2 = \int_{-\infty}^{\infty} \frac{(q-\lambda)^2}{\sqrt{\pi}\,\sigma}\,e^{-(q-\lambda)^2/\sigma^2}\,dq = \tfrac{1}{2}\,\sigma^2,$$

using the fact that

$$\int_{-\infty}^{\infty} x^2\,e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}.$$

Therefore, for every $\varepsilon > 0$, we can find a normalized function ψ_ε satisfying

$$\|\,Q\,\psi_\varepsilon - \lambda\,\psi_\varepsilon\,\| < \varepsilon, \qquad \|\,\psi_\varepsilon\,\| = 1.$$

We just set $\psi_\varepsilon = \widehat{\psi}_\sigma$ for any $\sigma < \sqrt{2}\,\varepsilon$.

A similar statement holds for the momentum operator $P = -\,\mathrm{i}\,\hbar\,\partial_q$: we just use the approximate normalized eigenfunctions

$$\widetilde{\psi}_\sigma(q) = \frac{1}{\pi^{1/4}\,\sigma^{1/2}}\,e^{-\,q^2/(2\,\sigma^2)}\,e^{\mathrm{i}\,\lambda\,q/\hbar},$$

for any $0 < \sigma < \hbar/(\sqrt{2}\,\varepsilon)$. The existence of such approximate eigenfunctions is one way of characterizing the continuous spectrum of such an operator; see below.

Now consider two different observables, represented by operators L and M. In order that they simultaneously have a definite value on a given state ψ is that ψ be a common eigenfunction of L and M:

$$L\,\psi = \lambda\,\psi, \qquad M\,\psi = \mu\,\psi.$$

In particular, this implies that

$$[L,M]\,\psi = 0,$$

i.e., the commutator between L and M vanishes on ψ. Conversely, if the operators commute, $[L,M] = 0$, then, assuming a purely discrete spectrum, they possess a complete set of common eigenfunctions, upon which they simultaneously can be measured with precision. In this sense, we say that commuting operators correspond to simultaneously measurable quantities.

For example, in the canonical quantization, the position and momentum operators (4.5), (4.6) satisfy the *canonical commutation relations*

$$[Q_j,Q_k] = 0 = [P_j,P_k], \qquad [Q_j,P_k] = \mathrm{i}\,\hbar\,\delta_{jk} = \begin{cases} \mathrm{i}\,\hbar, & j = k, \\[2mm] 0, & j \neq k. \end{cases} \tag{4.24}$$

Therefore, one can simultaneously measure all position operators or all momentum operators, or some positions and momenta as long as a position and its conjugate momentum are not both measured. In the latter case, we have the famous Heisenberg Uncertainty Relations, to be discussed next. For similar reasons, one cannot measure more than one angular momentum with certainty; see below for details.

The Uncertainty Principle

The justly famous *Heisenberg Uncertainty Principle* lies at the heart of quantum mechanics, and serve as the crux of many of its strange experimental manifestations and possible philosophical interpretations.

Theorem 4.3. *Let Q, P be operators satisfying*

$$[Q,P] = \mathrm{i}\,\hbar \tag{4.25}$$

Then the variations in the measurements of the two corresponding observables is restricted by the Uncertainty Relation

$$\Delta Q \cdot \Delta P \geq \tfrac{1}{2}\,\hbar. \tag{4.26}$$

Proof: By definition

$$\Delta Q = \sqrt{\langle (Q - \langle Q \rangle)^2 \rangle} = \sqrt{\langle \widehat{Q} \rangle^2}, \qquad \Delta P = \sqrt{\langle (P - \langle P \rangle)^2 \rangle} = \sqrt{\langle \widehat{P} \rangle^2},$$

where

$$\widehat{Q} = Q - \langle Q \rangle, \qquad \widehat{P} = P - \langle P \rangle.$$

Since $\langle Q \rangle, \langle P \rangle$ are scalars (or, rather, represent scalar multiples of the identity operator),

$$[\widehat{Q}, \widehat{P}] = i\hbar, \qquad \Delta \widehat{Q} = \Delta Q, \qquad \Delta \widehat{P} = \Delta P.$$

By the Cauchy–Schwarz inequality,

$$(\Delta Q)^2 (\Delta P)^2 = \langle \psi, \widehat{Q}^2 \psi \rangle \langle \psi, \widehat{P}^2 \psi \rangle \geq |\langle \psi, \widehat{Q}\,\widehat{P}\,\psi \rangle|^2.$$

Moreover,

$$\widehat{Q}\,\widehat{P} = \tfrac{1}{2}\big(\widehat{Q}\,\widehat{P} + \widehat{P}\,\widehat{Q}\big) + \tfrac{1}{2}[\widehat{Q}, \widehat{P}] = \widehat{R} + \tfrac{1}{2}i\hbar,$$

where $\widehat{R} = \tfrac{1}{2}\big(\widehat{Q}\,\widehat{P} + \widehat{P}\,\widehat{Q}\big)$ is the self-adjoint component of $\widehat{Q}\,\widehat{P}$. Therefore,

$$\langle \psi, \widehat{Q}\,\widehat{P}\,\psi \rangle = \langle \widehat{R} \rangle + \tfrac{1}{2}i\hbar$$

prescribes the real and imaginary parts of the left hand side, and hence

$$(\Delta Q)^2 (\Delta P)^2 \geq \langle \widehat{R} \rangle^2 + \tfrac{1}{4}\hbar^2,$$

which implies the uncertainty relation. Moreover, the uncertainty relation (4.26) will be an equality if and only if

    (a) $\widehat{Q}\,\psi = \lambda \widehat{P}\,\psi$, so that Cauchy–Schwarz is an equality, and

    (b) $0 = \langle \widehat{R} \rangle = (\lambda + \overline{\lambda})\langle \widehat{P}^2 \rangle$, so λ is purely imaginary.   *Q.E.D.*

Remark: By a slight modification of the above proof, one can generalize the Uncertainty Relation (4.26) to produce

$$[Q, P] = S \qquad \text{implies that} \qquad \Delta Q \cdot \Delta P \geq \tfrac{1}{2}\langle S \rangle. \tag{4.27}$$

5. Linear Operators and Their Spectrum.

<div align="right">

References: [**26**], [**45**], [**46**], [**50**].

</div>

Hilbert Space and Linear Operators

We work with a *complex Hilbert space* $\mathcal{E}$, i.e., a complete inner product space. For us, convergence means *convergence in norm*, i.e., $\psi_n \longrightarrow \psi$ means that for every $\varepsilon > 0$, there exists an N_ε such that $\|\psi_n - \psi\| < \varepsilon$ whenever $n > N_\varepsilon$. We will usually also assume that $\mathcal{E}$ is *separable*, meaning there exists a countable dense subset, which is equivalent to the

existence of a countable orthonormal basis $\varphi_1, \varphi_2, \varphi_3, \ldots$. If the basis is finite, then $\mathcal{E}$ is isomorphic to $\mathbb{C}^n$; if the basis is infinite, then $\mathcal{E}$ is isomorphic to the space

$$\ell^2 = \left\{ (z_1, z_2, z_3, \ldots) \;\middle|\; \sum_{j=1}^{\infty} |z_j| < \infty \right\} \tag{5.1}$$

consisting of all square summable sequences of complex numbers. For instance, $\mathrm{L}^2[0, 2\pi]$ is separable, since the complex exponentials $e^{\mathrm{i}nx}/\sqrt{2\pi}$ form a basis, by Fourier's Theorem, which implies that the Fourier series for an L^2 function converges in norm to that function.

A linear operator $L: \mathcal{E} \longrightarrow \mathcal{E}$ is *bounded* if there is a $c \geq 0$ such that

$$\| L\psi \| \leq c \| \psi \| \qquad \text{for all} \qquad \psi \in \mathcal{E}. \tag{5.2}$$

The smallest such c is called the *norm* of L, denoted $\| L \|$. It is not hard to see that $\| L \|$ defines a norm on the linear space of bounded operators on $\mathcal{E}$, and moreover

$$\| L \cdot M \| \leq \| L \| \, \| M \| \tag{5.3}$$

for any operators L, M. Bounded operators are the same as continuous linear operators. In finite dimensions, every linear transformation is continuous, and hence bounded; indeed,

$$\| L \| = \sup \{ \| L\psi \| \mid \| \psi \| = 1 \} < \infty,$$

where we use the fact that the unit ball in a finite-dimensional metric space is compact. This is no longer true in infinite dimensions. If L is only defined on a dense subspace, but satisfies (5.2) there, then it can easily be extended to a bounded operator on all of $\mathcal{E}$.

Unfortunately, most of the operators in quantum mechanics are unbounded. In fact, we have the following result:

Theorem 5.1. *Let Q, P be linear operators defined on a common dense subspace of a Hilbert space $\mathcal{E}$, which satisfy the canonical commutation relation $[Q, P] = \alpha$ where $\alpha \neq 0$. Then Q and P cannot both be bounded.*

Proof: By induction, starting with the canonical commutation relation, we find

$$Q P^n - P^n Q = n\alpha P^{n-1}. \tag{5.4}$$

First note that P (and Q) cannot be nilpotent, i.e., $P^n \neq 0$ for all n. Indeed, if $P^n = 0$ for some n, then (5.4) would show that $P^{n-1} = 0$ also, which is absurd. If both Q and P are bounded, then, using (5.3),

$$n \,|\, \alpha \,|\, \| P^{n-1} \| = \| n\alpha P^{n-1} \| \leq 2 \| Q \| \, \| P^n \| \leq 2 \| Q \| \, \| P \| \, \| P^{n-1} \|.$$

Now, by the first remark, $P^{n-1} \neq 0$, so $\| P^{n-1} \| \neq 0$. Therefore,

$$2 \| Q \| \, \| P \| \geq n \,|\, \alpha \,|, \qquad \text{for all} \qquad n = 0, 1, 2, 3, \ldots,$$

which is not possible if both are bounded. $\hfill$ *Q.E.D.*

Corollary 5.2. *There is no realization of the canonical commutation relations* (4.24) *on any finite-dimensional Hilbert space.*

Corollary 5.2 can also be easily proved by taking traces.

Even though we must in the end deal with unbounded operators, it is nevertheless extremely useful to have the easier theory of bounded operators under control. If $L: \mathcal{E} \longrightarrow \mathcal{E}$ is a bounded linear operator, then its kernel $\mathcal{K} = \ker L$ and range $\mathcal{R} = \operatorname{rng} L$ are subspaces of $\mathcal{E}$. In general, we can determine a unique inverse to L if and only if $\mathcal{K} = \{0\}$ is trivial, since then there will be a unique solution to the equation $L\varphi = \psi$ whenever $\psi \in \mathcal{R}$. In this case, the inverse of L will be the (possibly unbounded) linear operator $M: \mathcal{R} \longrightarrow \mathcal{E}$ defined on the range $\mathcal{R} \subset \mathcal{E}$, such that $\varphi = M\psi$ is the unique solution to $L\varphi = \psi$. In finite dimensions, L has an inverse if and only if $\mathcal{R} = \mathcal{E}$. In infinite dimensions, things are much more subtle, since $\mathcal{R}$ could be a subspace of $\mathcal{E}$ even when L has an inverse. For example, consider the shift operator

$$L(z_1, z_2, z_3, \dots) = (0, z_1, z_2, z_3, \dots) \tag{5.5}$$

on the space ℓ^2 of square summable sequences (5.1), which clearly has an inverse on its range

$$\mathcal{R} = \{(w_1, w_2, w_3, \dots) \mid w_1 = 0\} \subsetneqq \ell^2.$$

On the other hand, the inverse M can, in general, be extended to a bounded operator on all of $\mathcal{E}$ if and only if $\mathcal{R} = \mathcal{E}$.

The Spectral Theorem

In finite dimensions, the Spectral Theorem asserts that every symmetric (Hermitian) matrix possesses a basis of eigenvectors, and hence can be diagonalized, [**44**]. In terms of the eigenvector basis, the associated linear transformation just looks like multiplication by the eigenvalues, i.e., acts as a diagonal matrix. The Spectral Theorem for bounded operators on Hilbert space asserts essentially the same result.

Theorem 5.3. *Let L be a bounded operator on Hilbert space $\mathcal{E}$. Then there is a finite-dimensional measure space (X, μ) and a unitary map $U: L^2(X, \mu) \overset{\sim}{\longrightarrow} \mathcal{E}$ such that $U^* T U$ is multiplication by a complex valued function $f \in L^\infty(X, \mu)$.*

If L is self-adjoint, then f is real. Thus, in its spectral representation, L is given by multiplication by the function f. It is not hard to see that the spectrum of a multiplication operator is the same as the *essential range* of f, defined by

$$\operatorname{ess\,rng} f = \left\{ \lambda \;\middle|\; \mu\{x \mid \lambda - \varepsilon < f(x) < \lambda + \varepsilon\} > 0 \quad \text{for all} \quad \varepsilon > 0 \right\} \tag{5.6}$$

since the inverse of $f - \lambda$ is just $1/(f - \lambda)$, which is bounded provided λ is not in the essential range of f. Just as any measure μ can be decomposed into a pure point part and a continuous part, so the spectrum of an operator decomposes into a discrete part (the eigenvalues) and a continuous part. We will have more to say on this later.

The Spectral Theorem 5.3 can also be formulated in terms of spectral projections. Given the set $\{x \mid f(x) \leq \lambda\}$, let $\chi_\lambda(x)$ denote its *characteristic function*, so that $\chi_\lambda(x) = 1$

if $f(x) \leq \lambda$ and is zero otherwise. Thus, $\chi_\lambda \in \mathrm{L}^\infty(X, \mu)$, and we can consider the corresponding bounded operator $E_\lambda = U\chi_\lambda U^*$ which is a projection on $\mathcal{E}$. (For a finite-dimensional Hilbert space, E_λ would be projection onto the subspace spanned by all eigenvectors corresponding to eigenvalues $\leq \lambda$.) Then,

$$L = \int \lambda \, dE_\lambda \tag{5.7}$$

where the integral is a Riemann-Stieltjes integral. Using this representation, we can develop a functional calculus of bounded, self-adjoint operators, so that if F is any bounded, Borel measurable function on $\mathbb{R}$, then

$$F(L) = \int F(\lambda) \, dE_\lambda \tag{5.8}$$

is the operator whose spectral representation is multiplication by the function $F \circ f$. For example,

$$E_\lambda = \chi_\lambda(L), \qquad L^n = \int \lambda^n \, dE_\lambda, \qquad L^{-1} = \int \frac{dE_\lambda}{\lambda},$$

the latter as long as 0 is not in the spectrum of L. (In fact, the functional calculus makes sense provided F is only defined and bounded on the spectrum of L.) Note that the functional calculus of operators immediately provides a proof of the existence of the unitary group $e^{\mathrm{i}tH}$, when H is a bounded self-adjoint Hamiltonian. We just set

$$e^{\mathrm{i}tH} = \int e^{\mathrm{i}t\lambda} \, dE_\lambda. \tag{5.9}$$

The most interesting operators in quantum mechanics are unbounded operators. For example, the momentum operator $-\mathrm{i}\hbar\partial_q$ does not act on all of L^2, only on the dense subspace of C^1 functions whose derivatives are also in L^2 (or, more generally, the Sobolev space H^1). Nevertheless, for certain operators, a version of the Spectral Theorem 5.3 still holds. For instance, the momentum operator can be transformed into a multiplication operator by using the *Fourier transform*

$$\mathcal{F}\colon \mathrm{L}^2(\mathbb{R}, dq) \longrightarrow \mathrm{L}^2(\mathbb{R}, dk) \qquad \mathcal{F}[f(q)] = \widehat{f}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\mathrm{i}kq} f(q) \, dq$$

so that

$$\mathcal{F}\left(-\mathrm{i}\hbar \frac{df}{dq}\right) = \hbar k \, \widehat{f}(k).$$

Thus the spectrum of the momentum operator is entirely continuous, consisting of the entire real line, with spectral measure dk. This means that any value of the momentum of a quantum particle is observable, with equal probability. Moreover, since the function k is not bounded, the momentum operator cannot be bounded. For a direct proof of this, let $\varphi(x)$ be any L^2 function which is also C^1 and whose derivative φ' is also L^2. Consider the functions

$$\psi_\sigma(x) = \sqrt{\sigma}\, \varphi(\sigma x) \qquad \text{for} \qquad \sigma > 0.$$

These all have the same norm since

$$\| \, \psi_\sigma \, \|^2 = \int \sigma \, | \, \varphi(\sigma \, x) \, |^2 \, dx = \int | \, \varphi(y) \, |^2 \, dy = \| \, \varphi \, \|^2.$$

However, their derivatives are

$$\psi_\sigma' = \sigma^{3/2} \, \varphi(\sigma \, x),$$

and have unbounded norm since

$$\| \, \psi_\sigma' \, \|^2 = \int \sigma^3 \, | \, \varphi'(\sigma \, x) \, |^2 \, dx = \sigma^2 \int | \, \varphi'(y) \, |^2 \, dy = \sigma^2 \, \| \, \varphi' \, \|^2.$$

Spectrum of Operators

In finite dimensions, the spectrum of an operator (matrix) L coincides with its set of eigenvalues. These are the complex numbers λ such that $L - \lambda \, \mathrm{I}$ is not invertible. In infinite dimensions, things are more delicate. Given a bounded operator and a complex number $\lambda \in \mathbb{C}$, the operator $L - \lambda \, \mathrm{I}$ could be invertible, with either bounded or unbounded inverse, or not invertible at all. We distinguish these possibilities with the following definitions.

Definition 5.4. Let L be a bounded linear operator on a Hilbert space $\mathcal{E}$. The *resolvent set* of L is defined as

$$\rho(L) = \big\{ \, \lambda \, \big| \, (L - \lambda \, \mathrm{I})^{-1} \text{ exists as a bounded operator} \, \big\}. \qquad (5.10)$$

The operator $R_\lambda = (\lambda \, \mathrm{I} - L)^{-1}$ is called the *resolvent* of L. The complement of the resolvent set

$$\sigma(L) = \mathbb{C} \setminus \rho(L) \qquad (5.11)$$

is called the *spectrum* of L.

Lemma 5.5. *If L is an operator with range $\mathcal{E}$ and L^{-1} is bounded, then for $\varepsilon \in \mathbb{C}$ sufficiently small, $L + \varepsilon \, \mathrm{I}$ also has a bounded inverse.*

Proof:

$$(L + \varepsilon \, \mathrm{I})^{-1} = L^{-1} \, (\mathrm{I} + \varepsilon \, L^{-1})^{-1} = L^{-1} \sum_{n=0}^{\infty} (-\varepsilon \, L^{-1})^n.$$

Since $\| \, \varepsilon \, L^{-1} \, \| = | \, \varepsilon \, | \, \| \, L^{-1} \, \| < 1$ for ε sufficiently small, the operator series converges, and defines $(L + \varepsilon \, \mathrm{I})^{-1}$ as a bounded operator.   *Q.E.D.*

Corollary 5.6. *The resolvent set of L is an open subset of the complex plane, and so the spectrum is a closed subset.*

If L is self-adjoint, its spectrum $\sigma(L) \subset \mathbb{R}$ is a closed subset of the real line. We further divide the spectrum into three disjoint pieces.

Definition 5.7. Let L be a bounded operator on the Hilbert space $\mathcal{E}$. The *point spectrum* of L is

$$\sigma_p(L) = \{\, \lambda \mid L - \lambda\,\mathrm{I} \text{ has no inverse}\,\} = \{\, \lambda \mid L\psi = \lambda\psi \text{ for some } 0 \neq \psi \in \mathcal{E}\,\}.$$

The (naïve) *continuous spectrum* of L is

$$\sigma_c(L) = \{\, \lambda \mid L - \lambda\,\mathrm{I} \text{ has an unbounded inverse whose domain is dense in } \mathcal{E}\,\}.$$

The *residual spectrum* of L is

$$\sigma_\rho(L) = \{\, \lambda \mid L - \lambda\,\mathrm{I} \text{ has an inverse whose domain is not dense in } \mathcal{E}\,\}.$$

Thus, the spectrum is the disjoint union

$$\sigma(L) = \sigma_p(L) \ \vee\ \sigma_c(L) \ \vee\ \sigma_\rho(L) \tag{5.12}$$

of the point, continuous, and residual spectra.

Remark: Later, we will introduce a more sophisticated definition of the continuous spectrum, which will allow a number to be in both the point and continuous spectrum. As it stands, if we take the direct sum of two operators $L \oplus M$ such that λ is in the point spectrum of L and in the continuous spectrum of M, then, by the current definition, it ends up just in the point spectrum of the sum, which is not really correct.

Proposition 5.8. *If L is a self-adjoint operator, then L has no residual spectrum.*

Proof: Choose $0 \neq \varphi \in \mathcal{E}$ which is orthogonal to the domain of the inverse of $L - \lambda\,\mathrm{I}$, which is the same as the range of $L - \lambda\,\mathrm{I}$. Thus,

$$0 = \langle\, \varphi\,,(L - \lambda\,\mathrm{I})\,\psi\,\rangle = \langle\,(L - \lambda\,\mathrm{I})\,\varphi\,,\psi\,\rangle \quad \text{for all} \quad \psi \in \mathcal{E}.$$

This implies that $(L - \lambda\,\mathrm{I})\,\varphi = 0$, and hence φ is an eigenfunction of L which means that λ is in the point spectrum of L. But the three subsets are disjoint, so λ couldn't have been in the residual spectrum. *Q.E.D.*

As above, we define an *approximate eigenfunction* for the operator L to be a sequence of elements $\psi_1, \psi_2, \psi_3, \ldots \in \mathcal{E}$ satisfying

$$\|\psi_n\| = 1, \qquad \|L\psi_n - \lambda\psi_n\| \longrightarrow 0 \qquad \text{as} \qquad n \longrightarrow \infty, \tag{5.13}$$

for some $\lambda \in \mathbb{C}$.

Proposition 5.9. *Let L be a bounded linear operator on a Hilbert space $\mathcal{E}$ with no residual spectrum. Then L has an approximate eigenfunction corresponding to the complex number λ if and only if λ is in the spectrum of L.*

Proof: Let

$$A = L - \lambda\,\mathrm{I}, \qquad \chi_n = A\psi_n = L\psi_n - \lambda\psi_n.$$

If $\lambda \notin \sigma(L)$, then $B = A^{-1}$ is bounded, hence

$$1 = \|\psi_n\| = \|B\chi_n\| \leq \|B\|\,\|\chi_n\|.$$

Thus, we cannot have $\| \chi_n \| \longrightarrow 0$, and so there are no approximate eigenfunctions associated with λ.

Conversely, if λ is in the point spectrum, we just choose $\psi_n = \psi$ to be the associated eigenfunction for all n, so that $\| \chi_n \| = 0$ for all n. If λ is in the continuous spectrum, let χ_n be elements of the domain of $B = A^{-1}$ such that $\| \chi_n \| = 1$ and $\| B \chi_n \| \longrightarrow \infty$. Now, setting $\psi_n = \| \chi_n \|^{-1} B \chi_n$, we find

$$A \psi_n = \| \chi_n \|^{-1} \chi_n \longrightarrow 0,$$

so the sequence ψ_n defines an approximate eigenfunction. $\hspace{2cm}$ Q.E.D.

Remark: A point in the residual spectrum may or may not have an approximate eigenfunction. For instance, consider the shift operator (5.5). One finds that 0 belongs to its residual spectrum, but $\| L \psi \| = \| \psi \|$ for all $\psi \in \ell^2$.

Finally, we define the genuine continuous spectrum. Given a bounded operator L on the Hilbert space $\mathcal{E}$, let $\mathcal{E}_p \subset \mathcal{E}$ be the subspace spanned by the eigenfunctions of L. Note that $\mathcal{E}_p$ is invariant under L . If L is self-adjoint, its orthogonal complement $\mathcal{E}_c = \mathcal{E}_p^\perp$ is also invariant. These two subspace correspond to the point and genuine continuous spectrum respectively. We define the *continuous spectrum* $\widehat{\sigma}_c(L)$ to be the set of all λ for which there exist approximate eigenfunctions orthogonal to $\mathcal{E}_p$, i.e., $\psi_n \in \mathcal{E}_c$ for all n.

The most interesting operators in quantum mechanics are unbounded operators. For example, the momentum operator $- \mathrm{i} \hbar \partial_q$ does not act on all of L^2, only on the dense subspace of C^1 functions whose derivatives are also in L^2 (or, more generally, on the Sobolev space H^1). Let $\mathcal{E}$ be a Hilbert space, and let L be a linear operator from a subspace $\mathcal{D}(L) \subset \mathcal{E}$ into $\mathcal{E}$. In general, the domain $\mathcal{D}(L)$ is not closed, but we do assume that it is dense in $\mathcal{E}$. An operator $\widehat{L}$ is called an *extension* of L, written $L \subseteq \widehat{L}$, if $\mathcal{D}(L) \subset \mathcal{D}(\widehat{L})$ and $\widehat{L} = L$ on $\mathcal{D}(L)$. The operator L is called *closed* if its graph $\{ (\varphi, L \varphi) \, | \, \varphi \in \mathcal{D}(L) \}$ is a closed subset of $\mathcal{E} \times \mathcal{E}$. We call L *closable* if the closure of its graph is the graph of an operator, denoted $\overline{L}$, which is of course an extension of L. This is equivalent to the condition that if $\varphi_n \in \mathcal{D}(L)$ with $\varphi_n \longrightarrow 0$ and $L \varphi_n \longrightarrow \psi$, then $\psi = 0$.

The *adjoint* L^* of a densely defined linear operator is the linear operator defined as follows. Its domain is given by

$$\mathcal{D}(L^*) = \{ \, \varphi \, | \quad \text{there exists } \chi \in \mathcal{E} \text{ such that } \langle L \psi , \varphi \rangle = \langle \psi , \chi \rangle \text{ for all } \psi \in \mathcal{D}(L) \, \} .$$

For $\varphi \in \mathcal{D}(L^*)$, we set $L^*(\varphi) = \chi$ as defined above, verifying that it is uniquely specified and linear.

The linear operator L is *symmetric* if $L \subseteq L^*$. It is *self-adjoint* if $L = L^*$. It is *essentially self-adjoint* if $\overline{L} = L^*$, or, equivalently, $L^{**} = L^*$. The Spectral Theorem 5.3 only holds for self-adjoint operators, and only self-adjoint operators may be exponentiated to give one-parameter unitary groups.

The condition of self-adjointness is rather subtle. It depends not only on the operator being symmetric (formally self-adjoint), but also very much on the domain of definition, which in turn includes the boundary conditions we impose.

Example 5.10. Consider the momentum operator $P = -i \partial_x$. (We set $\hbar = 1$ for simplicity.) First consider a system on a finite interval $[a, b]$, so the Hilbert space is $\mathcal{E} = L^2 = L^2[a, b]$. Let $\mathcal{A} = \mathcal{A}[a, b]$ denote the set of *absolutely continuous* functions on $[a, b]$, i.e., functions ψ of the form

$$\psi(x) = \psi(a) + \int_a^x \varphi(y)\, dy \qquad \text{with} \qquad \varphi \in L^1, \tag{5.14}$$

where $\varphi = \psi'$ almost everywhere. These are the functions one can integrate by parts. Consider the domain

$$\mathcal{D}(P) = \left\{ \psi \mid \psi \in \mathcal{A},\ \psi' \in L^2,\ \psi(a) = \psi(b) = 0 \right\}.$$

The boundary conditions are to ensure that the operator P is symmetric. Indeed, if $\varphi, \psi \in \mathcal{D}(P)$, then integration by parts implies

$$\langle\, \varphi\,, P\psi \,\rangle = \int_a^b \varphi(x)\, \overline{(-i)\, \psi'(x)}\, dx = \left. i\, \varphi(x)\, \overline{\psi}(x)\right|_{x=a}^b + \int_a^b (-i)\, \varphi'(x)\, \overline{\psi}(x)\, dx = \langle\, P\varphi\,, \psi\,\rangle.$$

However, the boundary conditions are too strong for self-adjointness. Clearly the above computation does not depend on φ vanishing at the endpoints, although we do need φ to still be absolutely continuous with L^2 derivative. Therefore the adjoint P^* has domain

$$\mathcal{D}(P^*) = \left\{ \varphi \mid \varphi \in \mathcal{A},\ \varphi' \in L^2 \right\}.$$

and $P \subsetneq P^*$, so P is not self-adjoint on the above domain. Moreover, if we compute the adjoint P^{**} of P^* we find that we must reimpose the boundary conditions, so $P^{**} = P$, and hence P is closed. Thus P is a closed, symmetric, but not self-adjoint operator. Its adjoint P^* is a proper extension of P, but is not symmetric since $P^{**} \subsetneq P^*$.

However, it is possible to extend P to be self-adjoint. For instance, extending P to $\widehat{P}$, defined on the domain of periodic functions

$$\mathcal{D}(\widehat{P}) = \left\{ \psi \mid \psi \in \mathcal{A},\ \psi' \in L^2,\ \psi(a) = \psi(b) \right\}$$

will do. Indeed, the same integration by parts argument will be valid if and only if φ is also periodic, so in this case $\widehat{P} = \widehat{P}^*$. However, this is not the only way to extend P. We can also apply the same argument for

$$\mathcal{D}(P_\alpha) = \left\{ \psi \mid \psi \in \mathcal{A},\ \psi' \in L^2,\ \psi(a) = \alpha\, \psi(b) \right\}$$

for any complex number of modulus $|\alpha| = 1$. Thus, P has an entire "circle" of different possible self-adjoint extensions, $P \subsetneq P_\alpha \subsetneq P^*$, but no maximal self-adjoint extension.

What do these different self-adjoint extensions mean physically? If we take P to be the Hamiltonian, then the solutions of Schrödinger's equation[†]

$$i\, \psi_t = i\, \psi_x$$

---

are just translated waves
$$\psi(t, x) = f(x - t). \tag{5.15}$$

On a finite interval, a localized wave packet will, for small t, translate unchanged. However, after it reaches the endpoint 1, since the norm must be preserved, it has to reappear at the other endpoint 0, i.e., the interval should be viewed as bent into a circle. The different self-adjoint extensions of P then tell how the wave reappears, with α endowing it with an extra phase factor. Therefore, different self-adjoint extensions lead to different physics.

Next consider the same operator, but on the half line $[0, \infty)$. We have the domain

$$\mathcal{D}(P) = \left\{ \psi \mid \psi \in \mathcal{A}, \ \psi' \in \mathrm{L}^2, \ \psi(0) = 0 \right\}.$$

The integration by parts argument proceeds as before:

$$\langle \varphi, P\psi \rangle = \int_0^\infty \varphi(x) \overline{(-\mathrm{i}) \psi'(x)} \, dx = \left. \mathrm{i} \varphi(x) \overline{\psi}(x) \right|_0^\infty + \int_0^\infty (-\mathrm{i}) \varphi'(x) \overline{\psi}(x) \, dx = \langle P\varphi, \psi \rangle.$$

Again, the boundary conditions are too strong for self-adjointness, and

$$D(P^*) = \left\{ \varphi \mid \varphi \in \mathcal{A}, \ \varphi' \in \mathrm{L}^2 \right\}.$$

Then $P \subsetneq P^*$, so P is a closed, symmetric, but not self-adjoint operator. Its adjoint P^* is a proper extension of P, but is not symmetric since $P^{**} \subsetneq P^*$. However, in this case there is no possible way to extend P to be self-adjoint. (This will follow from Weyl's Theorem 5.11 on deficiency indices stated below.) In this case, the solutions to the Schrödinger equation run into problems when they hit the endpoint $x = 0$, and there is nowhere for them to go while simultaneously preserving the norm.

Finally, if we look at P on the entire real line, then no boundary conditions are needed, and P is self-adjoint on its natural domain of definition

$$\mathcal{D}(P) = \left\{ \psi \mid \psi \in \mathcal{A}, \ \psi' \in \mathrm{L}^2(\mathbb{R}) \right\}.$$

Now there are no endpoints to worry about, and the corresponding solutions to the Schrödinger equation are just translations (5.15) of the initial state $\psi(0, x) = f(x)$.

There is a useful criterion for an symmetric operator to admit self-adjoint extensions, due to Weyl, which we state without proof.

Theorem 5.11. *Let L be a symmetric operator on a Hilbert space $\mathcal{E}$. Then the following three conditions are equivalent:*

  (*i*) *L is essentially self-adjoint;*

  (*ii*) *$\ker(L^* \pm \mathrm{i}) = \{0\}$;*

  (*iii*) *$\mathrm{rng}\,(L \pm \mathrm{i})$ is dense in $\mathcal{E}$.*

More generally, we define the deficiency indices *of a closed symmetric operator to be*

$$m = \dim \ker(L^* + \mathrm{i}), \qquad n = \dim \ker(L^* - \mathrm{i}). \tag{5.16}$$

Then L has a self-adjoint extension if and only if $m = n$, in which case there is a (real) m^2–parameter family of different possible extensions corresponding to different possible isometric mappings of $\ker(L^* - \mathrm{i})$ onto $\ker(L^* + \mathrm{i})$.

For instance, the momentum operator P has deficiency indices $(1, 1)$ on the finite interval, since $e^{\pm x}$ lie in the kernel of $P^* \pm i$. This results in the 1-parameter family P_α of self-adjoint extensions which map e^x to αe^{-x} for $|\alpha| = 1$. On the half line, $P^* + i$ has no kernel (since e^x is not in L^2), whereas $P^* - i$ still has e^{-x} in its kernel. Thus the deficiency indices are $(0, 1)$, and no self-adjoint extension is possible. On the entire line, neither operator has a kernel, so the deficiency indices are both 0, and P is self-adjoint.

Example 5.12. Consider the operator $L = -d^2/dx^2$ on the entire real line. A domain is given by

$$\mathcal{D}(L) = \left\{ \psi \mid \psi \in \mathrm{C}^1, \quad \psi'' \in \mathrm{L}^2(\mathbb{R}) \right\}.$$

(This can be weakened by admitting functions in the *Sobolev space* H^2, consisting of L^2 functions whose first and second (weak) derivatives are also in L^2.) Then L is self-adjoint on $\mathcal{D}(L)$ since

$$\langle \varphi, L\psi \rangle = \int_{-\infty}^{\infty} \varphi(x) \left[\overline{-\psi''(x)} \right] dx = \left[\varphi'(x) \overline{\psi}(x) - \varphi(x) \overline{\psi}'(x) \right] \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \varphi''(x) \overline{\psi}(x) dx$$

$$= \int_{-\infty}^{\infty} \left[-\varphi''(x) \right] \overline{\psi}(x) dx = \langle L\varphi, \psi \rangle.$$

Note especially that the assumption that φ and ψ are C^1 is required so that no extraneous boundary terms show up inside the interval. The spectrum of L is continuous, and consists of the nonnegative real axis. Each $\lambda = k^2 > 0$ is doubly degenerate. Note that, for $\lambda < 0$, the solutions to $L\psi = \lambda\psi$ are real exponentials, so to obtain an element of the spectrum, we would need to have ψ decaying at both $\pm\infty$. We cannot piece them together to form an eigenfunction, e.g., by setting

$$\psi(x) = \begin{cases} e^{-kx}, & x > 0, \\ e^{kx}, & x < 0, \end{cases} \qquad \text{where} \qquad k > 0, \quad -k^2 = \lambda,$$

which, while certainly L^2, is no longer C^1, and lies outside the domain of definition of L. If we try to extend the domain of definition of L to include such piecewise smooth functions, we would no longer have a self-adjoint operator since the integration by parts would produce an extra contribution at $x = 0$. In the language of distributions, the second derivative of such a function is not, in fact in L^2, since it has a delta function discontinuity at $x = 0$ and hence is *not* a constant multiple of the function. Again, we see that domains of definition for self-adjointness are extremely important.

Here, one can explicitly construct the resolvent. For λ not on the nonnegative real axis, let $k^2 = -\lambda$ with $\mathrm{Re}\, k > 0$. Then, given $\varphi(x) \in \mathrm{L}^2$, the L^2 solution to

$$L\psi - \lambda\psi = -\psi'' - \lambda\psi = \varphi$$

is given by convolution with the Green's function:

$$\psi(x) = R_\lambda \varphi(x) = (L - \lambda I)^{-1} \varphi(x) = \int_{-\infty}^{\infty} \frac{e^{-k|x-y|}}{2k} \varphi(y)\, dy. \qquad (5.17)$$

Thus, for λ not on the nonnegative real axis, the resolvent is a bounded operator on L^2, and hence the spectrum is contained in the nonnegative real axis.

It is also instructive to consider the operator on the half line $[0, \infty)$. We begin with $L = - d^2/dx^2$ defined on the domain of C^∞ function of compact support in $(0, \infty)$. In this case, the deficiency indices are both 1 since the L^2 solutions of $-\psi'' = \pm\, \mathrm{i}\, \psi$ are multiples of

$$\psi_\pm(x) = \exp\left(\frac{-1 \pm \mathrm{i}}{\sqrt{2}}\, x \right).$$

It is not hard to see that the resulting one-parameter family of self-adjoint extensions of L have domain

$$\mathcal{D}(L_a) = \left\{\, \psi \,\mid\, \psi \in \mathrm{L}^2,\ \psi' \in \mathcal{A},\ \psi'' \in \mathrm{L}^2,\ \psi'(0) + \alpha\, \psi(0) = 0 \,\right\}.$$

where $\alpha \in \mathbb{R} \cup \{\infty\}$, with the boundary condition for $\alpha = \infty$ being $\psi(0) = 0$. Now, for $\alpha \neq \infty$, a bounded solution to the stationary Schrödinger equation in this domain is

$$e^{\mathrm{i}\, k\, x} + \beta\, e^{-\mathrm{i}\, k\, x},$$

where $\beta = (\mathrm{i}\, k + \alpha)/(\mathrm{i}\, k - \alpha)$, and $k = \sqrt{\lambda}$. This corresponds to a wave of momentum k coming in from ∞, and reflecting at $x = 0$ with the change of phase given by β. The case $\alpha = \infty$ corresponds to a hard wall potential, where the induced phase change for all momenta is $\beta = -1$.

So far we have been dealing with approximate eigenfunctions for these operators. However, these will not serve as legitimate state variables since they do not belong to L^2. To construct a bona fide wave function, we must sum up a collection of approximate eigenfunctions, weighted by their phase. Specifically, we look at the combination

$$\psi(x) = \int_{-\infty}^{\infty} \widehat{\psi}(k)\, e^{\mathrm{i}\, k\, x}\, dk,$$

corresponding to the weighting factor $\widehat{\psi}(k) \in \mathrm{L}^2$, which is just the Fourier transform of $\psi(x)$. Depending on the nature of the weighting factor, we will have different probability distributions. For instance, if $\widehat{\psi}(k)$ is concentrated near $k = k_0$, then $\psi(x)$ corresponds to a quantum particle whose momentum has, when measured, a value near $\hbar k_0$ with very high probability. This is because application of the momentum operator $P = -\mathrm{i}\, \hbar\, \partial_x$ to the approximate eigenfunction yields $\hbar k e^{\mathrm{i}\, k\, x}$, so its "expected value" is $\hbar k$. Thus, except for the factor of $\hbar$, we can identify the wave number k with the momentum of the corresponding particle. Similarly, if $\psi(x)$ is localized near x_0, then it corresponds to a quantum particle whose position measurements are concentrated near x_0. Note that, by the Uncertainty Principle, if the momentum is concentrated, the position is spread out over the entire line, and vice versa.

Now, consider the time evolution of such a wave function under the Schrödinger equation

$$\mathrm{i}\, \hbar\, \psi_t = -\, \hbar^2\, \psi_{xx},$$

where we have restored the factors of $\hbar$ so as to be better able to compare with the physics. The approximate eigenfunctions individually evolve according to

$$\psi(t, x, k) = e^{\mathrm{i}\,k\,x - \mathrm{i}\,\hbar\,k^2 t}, \tag{5.18}$$

which corresponds to a traveling periodic wave having wave velocity $c(k) = \hbar k$, which coincides with the momentum of the particle. Again, these solutions are not in L^2 and so are not physical. To construct a physical solution to the Schrödinger equation, we take a suitable superposition

$$\psi(t, x) = \int_{-\infty}^{\infty} \widehat{\psi}(k)\, e^{\mathrm{i}\,k\,x - \mathrm{i}\,\hbar\,k^2 t}\, dk,$$

corresponding to the weighting factor $\widehat{\psi}(k) \in \mathrm{L}^2$. Now, the behavior of this solution is less clear, although for specific weighting functions the integral can be evaluated explicitly, of course. For instance, if

$$\widehat{\psi}(k) = e^{-\alpha\,(k - k_0)^2}, \qquad \alpha > 0,$$

is a (un-normalized) Gaussian, then, completing the square in the integral yields

$$\psi(t, x) = \sqrt{\frac{\pi}{\alpha + \mathrm{i}\,\hbar\,t}}\; e^{\rho + \mathrm{i}\,\sigma},$$

where

$$\rho = -\frac{\alpha\,(x - 2\,\hbar\,k_0\,t)^2}{4\,\alpha^2 + \hbar^2\,t^2}, \qquad \sigma = -\frac{4\,\alpha^2\,k_0\,x - 4\,\alpha^2\,\hbar\,k_0^2\,t + \hbar\,x^2\,t}{4\,\alpha^2 + \hbar^2\,t^2}. \tag{5.19}$$

The corresponding probability density on position space is

$$|\,\psi(t, x)\,|^2\, dx = \frac{\pi\,e^{2\,\rho}}{\sqrt{\alpha^2 + \hbar^2\,t^2}}\; dx.$$

Note that the probability density is the highest where $\rho = 0$, i.e., where $x = 2\,\hbar\,t$. Therefore, the probability wave moves with combined velocity $2\,\hbar$, which is twice the velocity of the individual wave with wave number k_0.

Dispersion and Stationary Phase

References: **[28]**, **[40]**, **[57]**, **[60]**.

At first sight the preceding discussion may seem paradoxical (or even wrong), but is merely a manifestation of the well-known distinction between phase velocity and group velocity for a packet of waves in a dispersive medium. In general, a linear partial differential equation governing the propagation of waves is called *dispersive* if the individual Fourier modes move with different velocities, corresponding to complex exponential solutions of the form

$$\psi(t, x, k) = e^{\mathrm{i}\,k\,x - \mathrm{i}\,\omega\,t}, \tag{5.20}$$

where the wave number k and the temporal frequency ω are related according to the *dispersion relation* $\omega = \omega(k)$. For example in the case of the Schrödinger equation, in view of (5.18), the dispersion relation is $\omega = \hbar\,k^2$. The individual wave (5.20) moves with wave

speed or *phase velocity* $c_p = \omega(k)/k$, but a packet of waves with wave numbers concentrated near k moves with the *group velocity* $c_g = \omega'(k)$, which, except in trivial cases, e.g., the linear wave equation, is not the same. (It can be higher or lower, depending on the precise dispersion relation.) The energy in the system also travels with the group velocity, not the phase velocity. A familiar example is when you drop a rock into a still pond. The individual waves move with the phase velocity, but the group of waves corresponding to the disturbance moves with the group velocity. Thus, one observes (in the deep water case, where the phase velocity is approximately twice the group velocity) individual wave crests appearing at the back of the main disturbance, moving through the disturbance first growing in size and then decreasing, and finally disappearing at the front of the disturbance. (Try it!) In quantum mechanical contexts, a similar phenomena happens for the wave function solutions to Schrödinger's equation, except that the group velocity is now larger than the phase velocity, and so the waves appear at the front of the disturbance, and retreat to the back before disappearing.

Furthermore, there will be some spreading out of the wave due to dispersion. For the exponential $e^{-\alpha(x-a)^2}$, the larger α is, the more concentrated it is near $x = x_0$. In the present case, if α is small, then the initial wave function $\psi(x)$ is concentrated in position space near $x = 0$. (To introduce a phase shift to concentrate $\psi(x)$ near $x = \delta$, we would multiply the original weighting function by a complex exponential $e^{-i\delta}$.). However, as soon as $\hbar t$ gets to be of any negligible size, the corresponding coefficient $\alpha(t) = \alpha/(4\alpha^2 + \hbar^2 t^2)$, as prescribed in (5.19), becomes small, and the wave packet has spread out over a large region of position space. This reflects the fact that, by specifying the initial position of the particle rather precisely, the Uncertainty Principle implies that we have very little knowledge of the initial momentum, and so after a time the particle could have ended up almost anywhere. On the other hand, if α is large, so the initial momentum is well specified, although we have little knowledge of the initial position, then it remains reasonably localized in momentum space, but with a slight spreading as t increases. But the observed momentum of the particle is governed by the group velocity, not the phase velocity corresponding to its initial momentum!

For more general weighting functions, we can no longer evaluate the basic integral directly. However, our qualitative conclusions still hold. To derive these we need to analyze oscillatory integrals of the above and more general type. This can be done using the method of stationary phase, that dates back to Lord Kelvin in the nineteenth century.

Consider a superposition of the individual harmonic waves (5.20):

$$\psi(t, x) = \int_{-\infty}^{\infty} f(k)\, e^{\,i\,k\,x - i\,\omega(k)\,t}\, dk. \tag{5.21}$$

We are interested in the behavior for large t, x. The interesting limit is as $t \longrightarrow \infty$ with the speed $c = x/t$ held fixed, so we are looking at waves traveling with velocity c. Accordingly, we rewrite the integral as

$$\varphi(t) = \int_{-\infty}^{\infty} f(k)\, e^{-i\,\eta(k)\,t}\, dk, \qquad \text{where} \qquad \eta(k) = \omega(k) - c\,k. \tag{5.22}$$

According to Kelvin, the main contribution to the oscillatory integral (5.22) will be at the

stationary points of $\eta(k)$, i.e., those points for which

$$\eta'(k) = \omega'(k) - c = 0;$$

otherwise, the the complex exponential oscillates rapidly and makes no net contribution to the integral. Let us assume that there is just one such point k_0 for a given c. Note that, by the earlier discussion, these are precisely the points corresponding to the group velocity $c = \omega'(k)$ of the wave packet. Then we expand

$$\eta(k) = \eta(k_0) + \tfrac{1}{2}\,\eta''(k_0)\,(k - k_0)^2 + \;\cdots$$

near k_0, and assume $f(k_0) \neq 0$. The dominant contribution to the integral will be

$$\varphi(t) \equiv \int_{k_0-\varepsilon}^{k_0+\varepsilon} f(k_0)\exp\left(-\mathrm{i}\,t\left[\,\eta(k_0) + \tfrac{1}{2}\,\eta''(k_0)\,(k - k_0)^2\,\right]\right)dk$$

$$\equiv f(k_0)\,e^{-\mathrm{i}\,\eta(k_0)\,t}\int_{k_0-\varepsilon}^{k_0+\varepsilon}\exp\left(-\frac{\mathrm{i}}{2}\,\eta''(k_0)\,(k - k_0)^2\,t\right)dk$$

$$\equiv f(k_0)\,e^{-\mathrm{i}\,\eta(k_0)\,t}\int_{-\infty}^{\infty}\exp\left(-\frac{\mathrm{i}}{2}\,\eta''(k_0)\,(k - k_0)^2\,t\right)dk$$

$$\equiv f(k_0)\,e^{-\mathrm{i}\,\eta(k_0)\,t}e^{-\mathrm{i}\,\pi\,\mathrm{sign}\,\eta''(k_0)/4}\sqrt{\frac{2\,\pi}{t\,|\,\eta''(k_0)\,|}}\;.$$

The last integral (which is not absolutely convergent!) is evaluated by changing it into a real Gaussian integral via rotating the path of integration through $\pm\pi/4$, the sign depending on the sign of $\eta''(k_0)$. This can all be rigorously justified, [**40**].

Theorem 5.13. *Consider the integral*

$$\varphi(t) = \int_{-\infty}^{\infty} f(k)\,e^{\mathrm{i}\,\eta(k)t}\,dk. \tag{5.23}$$

where $\eta(k)$ is real, while $f(k)$ can be complex. Assume that
    (i) $\eta \in \mathrm{C}^4$;
    (ii) η' *has a finite number of nondegenerate zeros:* $\eta'(k_j) = 0,\ \eta''(k_j) \neq 0$;
    (iii) $f \in \mathrm{C}^3$ *with* f, f', f'' *bounded;*
    (iv) $(f/\eta')' \in \mathrm{L}^1$ *for* $|x| \gg 0$.
Then

$$\varphi(t) \;\sim\; \sum_j f(k_j)\exp\left[\frac{\mathrm{i}\,\pi\,\mathrm{sign}\,\eta''(k_j)}{4}\right]\sqrt{\frac{2\,\pi}{t\,|\,\eta''(k_j)\,|}}\,, \qquad t \longrightarrow \infty. \tag{5.24}$$

Proof: Using a partition of unity argument, we can assume that η has either one or no zeros in the support of f. If $\eta' \neq 0$ everywhere, then we can integrate by parts:

$$\int_{-\infty}^{\infty} f(k)\,e^{\mathrm{i}\,\eta(k)t}\,dk = \frac{1}{t}\int_{-\infty}^{\infty} g(k)\,e^{\mathrm{i}\,\eta(k)t}\,dk,$$

where $g(k) = (\mathrm{i}\, f(k)/\eta'(k))'$, which is L^1 by assumption. Thus the integral is $\mathrm{O}(t^{-1})$ at least. (In fact, if $f \in \mathrm{C}^\infty$ and of compact support, the integral is $\mathrm{O}(t^{-n})$ for any $n \in \mathbb{N}$.)

If $\eta'(k_0) = 0$ is the only zero of η' in the support of f, we proceed as follows. Without loss of generality, we assume $k_0 = 0$, $\eta(k_0) = 0$, and so $\eta(k) = \pm a\,k^2 + \mathrm{O}(k^3)$ near the critical point $k_0 = 0$, where $a = \frac{1}{2}\,|\,\eta''(0)\,| \neq 0$ by assumption. By a Lemma of Morse, we can introduce new coordinates $k = \zeta(u)$ near 0, such that

$$\eta(\zeta(u)) = \pm u^2.$$

If $\eta \in \mathrm{C}^k$, then it is not hard to see that ζ is at worst C^{k-1}. Then, changing variables, the integral becomes

$$\int_{-\infty}^{\infty} h(u)\, e^{\pm\mathrm{i}\,u^2 t}\, du$$

where $h(u) = f(\zeta(u))\,\zeta'(u)$ is bounded, with its first two derivatives bounded.

Lemma 5.14. *Suppose $h(u)$ is a C^2 function on $\mathbb{R}$ such that h, h', h'' are bounded. Then*

$$\int_{-\infty}^{\infty} h(u)\, e^{\pm\mathrm{i}\,u^2 t}\, du = h(0)\,\sqrt{\frac{\pi}{t}}\, e^{\pm\mathrm{i}\,\pi/4} + \mathrm{O}(t^{-3/2}), \qquad t \longrightarrow \infty. \qquad (5.25)$$

Proof: For brevity, we just do the case where the sign is positive. We first need to show that the integral exists. To prove that it converges uniformly in t for $t \neq 0$, let a, b be real, a large negative, b large positive. Integrating by parts, we find

$$\int_a^b h(u)\, e^{\mathrm{i}\,u^2 t}\, du = \frac{1}{2\,\mathrm{i}\,t} \int_a^b \frac{h(u)}{u}\, (e^{\mathrm{i}\,u^2 t})'\, du = \frac{1}{2\,\mathrm{i}\,t}\, \frac{h(u)}{u}\, e^{\mathrm{i}\,u^2 t}\, \Big|_a^b - \frac{1}{2\,\mathrm{i}\,t} \int_a^b \left(\frac{h(u)}{u}\right)'\, e^{\mathrm{i}\,u^2 t}\, du$$

$$= \left[\frac{1}{2\,\mathrm{i}\,t}\, \frac{h(u)}{u} - \frac{1}{(2\,\mathrm{i}\,t)^2 u} \left(\frac{h(u)}{u}\right)'\right] e^{\mathrm{i}\,u^2 t}\, \Big|_a^b + \frac{1}{(2\,\mathrm{i}\,t)^2} \int_a^b \left(\frac{1}{u} \left(\frac{h(u)}{u}\right)'\right)'\, e^{\mathrm{i}\,u^2 t}\, du.$$

Now

$$\left(\frac{h(u)}{u}\right)' = \frac{h'(u)}{u} - \frac{h(u)}{u^2} \qquad \left(\frac{1}{u} \left(\frac{h(u)}{u}\right)'\right)' = \frac{h''(u)}{u^2} - 3\,\frac{h'(u)}{u^3} + 3\,\frac{h(u)}{u^4}.$$

Therefore, provided h, h', h'' are bounded, the boundary terms go to zero as $a \to -\infty$, $b \to \infty$, and the final integral is absolutely convergent. Therefore, the original integral converges uniformly as $a \to -\infty$, $b \to \infty$.

The second step is to show that

$$\int_{-\infty}^{\infty} e^{\mathrm{i}\,u^2 t}\, du = \sqrt{\frac{\pi}{t}}\, e^{\pm\mathrm{i}\,\pi/4}. \qquad (5.26)$$

To do this, note that the integral

$$F(\lambda) = \int_{-\infty}^{\infty} e^{\lambda u^2}\, du$$

is convergent for $\operatorname{Re}\lambda \leq 0$ except for $\lambda = 0$, is continuous as a function of λ, and, moreover, is an analytic function of λ for $\operatorname{Re}\lambda < 0$. Moreover, for $0 > \lambda$ real, we know that

$$F(\lambda) = \sqrt{\frac{\pi}{-\lambda}}\ .$$

Therefore, to evaluate $F(\lambda)$ on the imaginary axis $\lambda = \pm\,\mathrm{i}\,t$, we must analytically continue the square root $+\sqrt{-\lambda}$ to the imaginary axis, which leads to the branch

$$\sqrt{\pm\,\mathrm{i}\,t} = \sqrt{t}\ e^{\mp\,\mathrm{i}\,\pi/4}.$$

This proves formula (5.26).

To prove the Lemma, we replace $h(u)$ by $h(u) - h(0)$, which accounts for the leading term in the asymptotic expansion. Therefore, we need only show that if $h(0) = 0$, then the integral can be bounded by a multiple of $t^{-3/2}$. We perform the same integration by parts as above, to deduce that in this case

$$\int_{-\infty}^{\infty} h(u)\,e^{\,\mathrm{i}\,u^2 t}\,du = -\frac{1}{2\,\mathrm{i}\,t}\int_{-\infty}^{\infty}\widehat{h}(u)\,e^{\,\mathrm{i}\,u^2 t}\,du, \qquad \text{where} \qquad \widehat{h}(u) = \left(\frac{h(u)}{u}\right)'$$

which, since $h(0) = 0$, is C^1. Iterating, we find that

$$\int_{-\infty}^{\infty}\widehat{h}(u)\,e^{\,\mathrm{i}\,u^2 t}\,du = \widehat{h}(0)\,\sqrt{\frac{\pi}{t}}\ e^{\pm i\pi/4} - \frac{1}{2\,\mathrm{i}\,t}\int_{-\infty}^{\infty}\widetilde{h}(u)\,e^{\,\mathrm{i}\,u^2 t}\,du,$$

where

$$\widetilde{h}(u) = \left(\frac{\widehat{h}(u) - \widehat{h}(0)}{u}\right)'.$$

As before, the last integral is absolutely convergent, so that it can be bounded by a constant K independent of t. Therefore,

$$\left|\int_{-\infty}^{\infty}\widehat{h}(u)\,e^{\,\mathrm{i}\,u^2 t}\,du\right| \leq \frac{|\,\widehat{h}(0)\,|\,\sqrt{\pi}}{2\,t^{3/2}} + \frac{K}{t^2} = \frac{|\,h''(0)\,|\,\sqrt{\pi}}{4\,t^{3/2}} + \frac{K}{t^2},$$

which establishes Lemma 5.14, and hence Theorem 5.13. $\hspace{2cm}$ Q.E.D.

6. One–Dimensional Problems.

References: [**26**], [**36**], [**45**], [**46**], [**50**].

Consider a one-dimensional Hamiltonian of the form

$$H(p,q) = \frac{p^2}{2\,m} + V(q), \tag{6.1}$$

corresponding to a particle of mass m in a potential force field $V(q)$. The corresponding stationary Schrödinger equation is

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi = \lambda \, \psi. \tag{6.2}$$

For simplicity, we choose units so that $\hbar^2 = 2m$; equivalently, rescale both $V(x)$ and λ by $2m/\hbar^2$ times themselves. Thus, we look at the equation

$$\psi'' + \left[\lambda - V(x) \right] \psi = 0 \tag{6.3}$$

on the entire line, so $x \in \mathbb{R}$.

Example 6.1. *Step potential*: Consider the case of potential in the form of a step function:

$$V(x) = \left\{ \begin{array}{ll} a, & x > 0, \\ b, & x < 0, \end{array} \right. \tag{6.4}$$

where we assume that $a < b$. We want a solution $\psi(x)$ to Schrödinger equation (6.2) which is (at least) C^1. There are three cases:

Case 1: If $\lambda < a$, then the solution is exponential in both regions. However, it is not hard to see that it cannot be C^1 and exponentially decaying at both $\pm\infty$. A similar statement holds if $\lambda = a$.

Case 2: If $a < \lambda < b$, then the solution is exponential for $x \longrightarrow -\infty$, and oscillatory for $x \longrightarrow +\infty$. Up to multiple, there is exactly one function which remains bounded for all x. Therefore, each such λ is in the continuous spectrum, with multiplicity 1. Similar remarks hold for $\lambda = b$.

Case 3: If $b < \lambda$, then the solution is oscillatory in both regions. Thus there are two bounded C^1 solutions for each such λ, which is hence in the continuous spectrum, with multiplicity 2.

Now we analyze the behavior of the quantum system, and contrast it with its classical counterpart. In Case 2, a classical particle will come in from the positive x axis with speed $k = \sqrt{\lambda - a}$, will rebound elastically at the barrier $x = 0$, and then travel back out to ∞ with the same speed.

The corresponding quantum solution takes the form

$$\psi(x) = \left\{ \begin{array}{ll} e^{-ikx} + R \, e^{ikx}, & x > 0, \\ T \, e^{\kappa x}, & x < 0, \end{array} \right.$$

where $k = \sqrt{b - \lambda}$. Because $\psi \in C^1$, we find

$$R = e^{i\varphi} = \frac{ik - \kappa}{ik + \kappa}, \qquad T = 1 + R.$$

The corresponding time-dependent solution has the form

$$\psi(t, x) = e^{-ikx - i\lambda t} + e^{ik(x+\delta) - i\lambda t} \qquad \text{for} \qquad x > 0,$$

which consists of an incident wave traveling with velocity λ/k, and a reflected wave traveling with the opposite velocity, but suffering a time delay δ experienced upon collision. Furthermore, for $x < 0$, the wave function $\psi(t, x) \neq 0$, and so there is a nonzero probability of finding the particle o the other side of the potential barrier. Of course, these solutions are not C^1 as they stand, and so, to have a physically meaningful statement, we must superimpose several solutions and apply the stationary phase approximation to the integrals. We leave it to the reader to verify that the statements we have made carry over to this approximation.

In Case 3, a classical particle will come in from the positive x axis with speed $k = \sqrt{\lambda - a}$, and will continue on to $-\infty$ with reduced speed $l = \sqrt{\lambda - b}$, or the reverse. The corresponding quantum solution takes the form

$$y(x) = \begin{cases} e^{-ikx} + R\,e^{ikx}, & x > 0, \\ T\,e^{ilx}, & x < 0, \end{cases}$$

where, because $\psi \in \mathrm{C}^1$, we find

$$R = \frac{k - l}{k + l}, \qquad T = 1 + R,$$

are real. The quantity R is called the *reflection coefficient*, and measures the probability that the particle will be reflected by the barrier, whereas T is called the *transmission coefficient*, and represents the probability that it will be transmitted. Note that $|T|^2$ measures the probability of the particle being transmitted though the barrier, while $|R|^2$ measures the probability of the particle being reflected. In all cases, there is a nonzero probability of the particle bouncing off the potential barrier.

Example 6.2. *Penetration of a square potential barrier or square well.*
Suppose

$$V(x) = \begin{cases} h, & 0 < x < a, \\ 0, & x < 0 \quad \text{or} \quad x > a. \end{cases} \tag{6.5}$$

Again, we seek a solution $\psi(x)$ which is C^1. There are two cases:

Case 1: If $h > 0$, we have a square potential barrier. For $\lambda < 0$, the quantum mechanical solution $\psi(x)$ is exponential all three regions. However, it is not hard to see that a C^1 solution is monotone, and hence the spectrum can only be positive. For $\lambda = k^2 > 0$, the solution is oscillatory and hence each such λ belongs to the continuous spectrum, with multiplicity 2.

Consider the solution of the form

$$\psi(x) = \begin{cases} e^{-ikx} + Re^{ikx}, & x > a, \\ Te^{-ikx}, & x < 0. \end{cases}$$

In the region $0 < x < a$ the solution will be either exponential or sinusoidal depending on whether λ is less than or greater than h. The transmission coefficient can be straightfor-

wardly calculated to be

$$|T|^2 = \begin{cases} \dfrac{\mu}{\mu + h^2 \sin^2(\kappa a)}, & \lambda > h, \\[3mm] \dfrac{\mu}{\mu + h^2 \sinh^2(\kappa a)}, & \lambda < h, \end{cases} \qquad \text{where} \qquad \begin{aligned} \kappa &= \sqrt{|\lambda - h|}, \\[2mm] \mu &= 4\lambda(\lambda - h). \end{aligned}$$

Also

$$|R|^2 + |T|^2 = 1,$$

so the probabilities of being transmitted or reflected add up to 1, as they should.

As before, we can contrast the behavior of a classical and quantum particle. The classical particle coming in from $+\infty$ will rebound off the barrier if its energy λ is less than the potential height h; otherwise it will be slowed down in the barrier, but traverses it and continues on to $-\infty$. In contrast, the quantum particle always has a nonzero probability of penetrating the barrier. As λ increases from 0 to h, the transmission coefficient increases monotonically from 0 to $\left(1 + \frac{1}{4}h\lambda^2\right)^{-1}$. This effect is known as the *tunnel effect* and plays an important role in radioactivity. For $\lambda > h$, complete transmission, $|T|^2 = 1$, occurs only for special values of the energy, namely when κa is a multiple of π. As the energy increases, the transmission probability oscillates between 1 and a minimum value on the order of $\mu(2\lambda - h)^{-2}$. The effect is particularly marked when the barrier is very high or thick, and when the kinetic energy $\lambda - h$ in the region of the barrier is small.

Case 2: If $h < 0$, we have a square potential well. For $\lambda = -k^2 < 0$, the solution is exponential in the regions $x < 0$, $x > a$. If $\lambda \leq h$, then it is also exponential in the well, and as before, any C^1 solution must be monotone, so it cannot be exponentially decaying at both ends. If $h < \lambda < 0$, then any bounded solution must, up to multiple, take the form

$$\psi(x) = \begin{cases} e^{-kx}, & x > a, \\ A\sin(\kappa x + \delta), & 0 < x < a, \\ Be^{kx}, & x < 0, \end{cases} \qquad \text{where} \qquad \kappa = \sqrt{\lambda - h},$$

and A and B are real. If ψ is to be C^1, we must have

$$A\sin\delta = B, \qquad A\kappa\cos\delta = Bk, \qquad A\sin(\kappa a + \delta) = B, \qquad A\kappa\cos(\kappa a + \delta) = -Bk.$$

We can determine A, B provided

$$\tan\delta = \frac{\kappa}{k} = -\tan(\kappa a + \delta).$$

We have

$$\sin^2\delta = \frac{\tan^2\delta}{1 + \tan^2\delta} = \frac{\kappa^2}{\kappa^2 + k^2} = \frac{h + \lambda}{h} = \frac{\kappa^2}{h}$$

Eliminating the phase shift δ,

$$\kappa a + 2\sin^{-1}\frac{\kappa}{\sqrt{h}} = n\pi, \tag{6.6}$$

for some integer n. There are a finite number $m \geq 0$ of solutions κ, n to (6.6), corresponding to a finite number of negative eigenvalues $\lambda_1 < \lambda_2 < \ldots < \lambda_m < 0$ of the Schrödinger equation — m may be zero if the well is too shallow — and so to a finite number of negative energy bound states of the quantum mechanical problem. In the classical picture of a potential well, there are bound states corresponding to all values of the energy greater than the minimum of the well. Thus, the Schrödinger equation has served to quantize the bound states of the classical system into discrete energy levels, while retaining the continuous character of the non-bound or scattering states.

Consider a scattering solution

$$
\psi(x) = \begin{cases} e^{-ikx} + R\,e^{ikx}, & x > a, \\ T\,e^{-ikx}, & x < 0, \end{cases}
$$

with reflection coefficient R and transmission coefficient T. For $\lambda > 0$, the transmission and reflection coefficients can be straightforwardly calculated to be

$$
|T|^2 = \frac{\mu}{\mu + h^2 \sin^2 ka}, \qquad |R|^2 = 1 - |T|^2, \qquad \text{where} \qquad \begin{aligned} \kappa &= \sqrt{\lambda - h}, \\ \mu &= 4\lambda(\lambda - h). \end{aligned}
$$

Note that, in contrast with a classical particle, a quantum particle usually has a non-zero probability of being reflected by the potential well. Exceptions occur when $\sin \kappa a = 0$, at which points, the reflection coefficient is zero and the particle is completely transmitted. Thus, there is a resonance effect in the quantum probabilities of transmission/reflection.

The Harmonic Oscillator

The classical *harmonic oscillator* has energy

$$
H(p,q) = \frac{1}{2m}(p^2 + m^2\omega^2 q^2), \tag{6.7}
$$

leading to the Hamiltonian system

$$
\frac{dp}{dt} = -m\,\omega^2 q, \qquad \frac{dq}{dt} = \frac{p}{m}, \tag{6.8}
$$

with is a linear system of first order ordinary differential equations. The resulting motion is oscillatory of frequency ω.

The corresponding quantum mechanical Hamiltonian operator takes the same form, but where q and p are operators satisfying the canonical commutation relation

$$
[q, p] = i\,\hbar.
$$

Let us rescale to eliminate the inessential constants:

$$
Q = \sqrt{\frac{m\,\omega}{\hbar}}\, q, \qquad P = \frac{1}{\sqrt{m\,\hbar\,\omega}}\, p, \qquad \mathcal{H} = \frac{1}{\sqrt{\hbar\,\omega}} H,
$$

so that

$$
\mathcal{H} = \tfrac{1}{2}(P^2 + Q^2),
$$

where Q, P are operators satisfying

$$[Q, P] = \mathrm{i}.$$

In the Schrödinger representation

$$Q = x, \qquad P = -\mathrm{i}\,\frac{d}{dx},$$

and the stationary Schrödinger equation is

$$\frac{1}{2}\left(-\frac{d^2}{dx^2} + x^2\right)\psi = \lambda\,\psi. \tag{6.9}$$

As we will see below, the solutions to this eigenvalue problem are expressed in terms of Hermite polynomials. It is convenient to proceed a little more abstractly as follows:

Define the "creation" and "annihilation" operators

$$A^* = \frac{1}{\sqrt{2}}\,(Q - \mathrm{i}\,P), \qquad A = \frac{1}{\sqrt{2}}\,(Q + \mathrm{i}\,P), \tag{6.10}$$

so

$$[A, A^*] = 1,$$

and

$$\mathcal{H} = N + \tfrac{1}{2}, \qquad \text{where} \qquad N = A^*A.$$

Note that eigenfunctions of N are eigenfunctions of $\mathcal{H}$, but the eigenvalues are shifted by $\frac{1}{2}$. Furthermore, note

$$A A^* = N + 1, \qquad N A = A(N - 1), \qquad N A^* = A^*(N + 1).$$

Now suppose ψ_λ is an eigenfunction of N corresponding to the eigenvalue λ. Then $\psi_{\lambda-1} = A\psi_\lambda$ (if not zero) is an eigenfunction of N with eigenvalue $\lambda - 1$, and $\psi_{\lambda+1} = A^*\psi_\lambda$ (if not zero) is an eigenfunction of N corresponding to the eigenvalue $\lambda + 1$:

$$N\psi_{\lambda-1} = NA\psi_\lambda = A(N-1)\psi_\lambda = (\lambda - 1)A\psi_\lambda = (\lambda - 1)\psi_{\lambda-1},$$
$$N\psi_{\lambda+1} = NA^*\psi_\lambda = A^*(N+1)\psi_\lambda = (\lambda + 1)A^*\psi_\lambda = (\lambda + 1)\psi_{\lambda+1}.$$

Moreover,

$$\|\psi_{\lambda-1}\|^2 = \|A\psi_\lambda\|^2 = \langle A\psi_\lambda, A\psi_\lambda \rangle = \langle \psi_\lambda, A^*A\psi_\lambda \rangle = \langle \psi_\lambda, N\psi_\lambda \rangle = \lambda\|\psi_\lambda\|^2$$

which implies that $\lambda \geq 0$, with $A\psi_\lambda = 0$ if and only if $\lambda = 0$. Applying the same argument to $\psi_{\lambda-1}$, with $\psi_{\lambda-2} = A\psi_{\lambda-1}$, we deduce that, if $\lambda \neq 0$, then $\lambda \geq 1$. Continuing in this fashion, we conclude that every eigenvalue of N is a nonnegative integer, and, moreover, every nonnegative integer belongs to the spectrum, since

$$\|\psi_{\lambda+1}\|^2 = \langle A^*\psi_\lambda, A^*\psi_\lambda \rangle = \langle \psi_\lambda, AA^*\psi_\lambda \rangle = \langle \psi_\lambda, (N+1)\psi_\lambda \rangle = (\lambda + 1)\|\psi_\lambda\|^2$$

is never zero for $\lambda \geq 0$. We normalize, so that ψ_n is now the normalized eigenfunction corresponding to the eigenvalue $0 \leq n \in \mathbb{N}$. Then

$$A\psi_n = \sqrt{n-1}\,\psi_{n-1}, \qquad A^*\psi_n = \sqrt{n+1}\,\psi_{n+1}. \tag{6.11}$$

If one interprets the state ψ_n as having n quanta, then the operator A^* "creates" an additional quantum, while the operator A^* "annihilates" one, hence the names *creation* and *annihilation operators*. One can also interpret the Hamiltonian as describing a system of indistinguishable particles, all in the same dynamical state, whose energy is $\hbar\omega$. In this case ψ_n represents the state having n particles, with ψ_0 being the vacuum state. In this case, the operator N represents the number of particles, and the operators A^* and A "create" and "destroy" particles.

In the Schrödinger representation,

$$A^* = \frac{1}{\sqrt{2}}\left(-\frac{d}{dx} + x\right), \qquad A^* = \frac{1}{\sqrt{2}}\left(\frac{d}{dx} + x\right). \tag{6.12}$$

The vacuum ψ_0 must satisfy $A\psi_0 = 0$, i.e.,

$$\frac{d\psi_0}{dx} = -x\psi_0, \qquad \text{hence} \qquad \psi_0(x) = \frac{1}{\pi^{1/4}}\,e^{-x^2/2}$$

is a normalized Gaussian. Then

$$\psi_n(x) = \frac{1}{\pi^{1/4}\sqrt{n!}}\,(A^*)^n\,e^{-x^2/2} = \frac{1}{2^{n/2}\,\pi^{1/4}\sqrt{n!}}\,H_n(x)e^{-x^2/2}, \tag{6.13}$$

where H_n is the nth *Hermite polynomial*

$$H_n(x) = (-1)^n\,e^{x^2}\frac{d^n}{dx^n}e^{-x^2}. \tag{6.14}$$

The eigenfunctions (6.13) form a complete orthonormal basis in L^2. The operator N, and hence also $\mathcal{H}$, is essentially self-adjoint, and can be extended to a self-adjoint operator on a suitable subdomain of L^2.

Scattering and Inverse Scattering

References: [**1**], [**14**], [**21**], [**48**].

We return to the *one-dimensional Schrödinger equation*

$$\psi'' + \left[\,\lambda - V(x)\,\right]\psi = 0. \tag{6.15}$$

The potential $V(x)$ is required to tend to 0 sufficiently rapidly as $|x| \to \infty$, in order that

$$\int_{-\infty}^{\infty} (1 + x^2)\,|V(x)|\,dx < \infty. \tag{6.16}$$

We are interested in the quantum scattering problem for this equation. Thus, we will look at solutions having the asymptotic form

$$\psi(x) \ \sim \ \begin{cases} e^{-ikx} + R(k)\,e^{ikx}, & x \longrightarrow +\infty \\ T(k)\,e^{-ikx}, & x \longrightarrow -\infty, \end{cases} \tag{6.17}$$

where $k^2 = \lambda$. As in Example 6.1, $R(k)$ is known as the *reflection coefficient* and $T(k)$ the *transmission coefficient*. Physically, $|R(k)|^2$ gives the probability of a particle with momentum k reflecting off of the potential $V(x)$, while $|T(k)|^2$ gives the probability of the particle passing through the potential.

The asymptotic form (6.17)is to be expected because of the decay of the potential at $\pm\infty$. The rigorous result is the following.

Theorem 6.3. *Assume that the potential $V(x)$ satisfies (6.16). Given $k \in \mathbb{C}$ with $\operatorname{Im} k \geq 0$, there exist unique solutions $\psi(x,k)$, $\varphi(x,k)$ to the Schrödinger equation*

$$\psi'' + \left[\, k^2 - V(x)\,\right]\psi = 0$$

satisfying the uniform asymptotic estimates

$$\psi(x,k) \sim e^{\,\mathrm{i}\,k\,x}\left(\,1 + \mathrm{o}(1)\,\right), \qquad \psi_x(x,k) \sim e^{\,\mathrm{i}\,k\,x}\left(\,\mathrm{i}\,k + \mathrm{o}(1)\,\right), \qquad x \longrightarrow +\infty,$$
$$\varphi(x,k) \sim e^{-\,\mathrm{i}\,k\,x}\left(\,1 + \mathrm{o}(1)\,\right), \qquad \varphi_x(x,k) \sim e^{-\,\mathrm{i}\,k\,x}\left(-\,\mathrm{i}\,k + \mathrm{o}(1)\,\right), \qquad x \longrightarrow -\infty.$$

Moreover, for each fixed x, the solutions ψ, φ, and their x derivatives are analytic functions of k for $\operatorname{Im} k > 0$ and continuous for $\operatorname{Im} k \geq 0$.

Furthermore, for $\operatorname{Im} k \leq 0$, the complex conjugate solutions

$$\overline{\psi}(x,k) = \psi(x,-k) = \overline{\psi(x,\overline{k})}, \qquad \overline{\varphi}(x,k) = \varphi(x,-k) = \overline{\varphi(x,\overline{k})},$$

satisfy the alternative asymptotic estimates

$$\overline{\psi}(x,k) \sim e^{-\,\mathrm{i}\,k\,x}\left(\,1 + \mathrm{o}(1)\,\right), \qquad \overline{\psi}_x(x,k) \sim e^{-\,\mathrm{i}\,k\,x}\left(-\,\mathrm{i}\,k + \mathrm{o}(1)\,\right), \qquad x \longrightarrow +\infty,$$
$$\overline{\varphi}(x,k) \sim e^{\,\mathrm{i}\,k\,x}\left(\,1 + \mathrm{o}(1)\,\right), \qquad \overline{\varphi}_x(x,k) \sim e^{\,\mathrm{i}\,k\,x}\left(\,\mathrm{i}\,k + \mathrm{o}(1)\,\right), \qquad x \longrightarrow -\infty,$$

and are analytic for $\operatorname{Im} k < 0$ and continuous for $\operatorname{Im} k \leq 0$. For $k = 0$, one also requires solutions which asymptote to x at $+\infty$ or at $-\infty$. The solutions $\psi, \varphi, \overline{\psi}, \overline{\varphi}$ are known as the Jost solutions *for the problem.*

Proof: This is fairly standard result from the theory of ordinary differential equations. We replace the Schrödinger equation by the linear integral equation

$$\psi(x,k) = e^{\,\mathrm{i}\,k\,x} + \frac{1}{k}\int_x^\infty V(y)\,\psi(y,k)\sin k\,(y-x)\,dy \tag{6.18}$$

which is solved by the method of successive approximation. In the course of the proof, we establish the important estimate

$$\left|\,e^{-\,\mathrm{i}\,k\,x}\,\psi(x,k)\,\right| \leq \exp\left(\frac{1}{k}\int_x^\infty |\,V(y)\,|\,dy\right). \tag{6.19}$$

This proves the existence of the Jost solution ψ, and the uniqueness also follows from the standard theory of ordinary differential equations. The corresponding result for φ is similar; details appear in the references, especially [**21**]. If we substitute (6.19) into the integral equation (6.18), we deduce that

$$e^{-\,\mathrm{i}\,k\,x}\,\psi(x,k) = 1 + \frac{1}{2\,\mathrm{i}\,k}\int_x^\infty V(y)\,dy + \mathrm{O}\left(\,|\,k\,|^{-2}\,\right). \tag{6.20}$$

Similarly, we find

$$e^{\,\mathrm{i}\,k\,x}\,\varphi(x,k) = 1 - \frac{1}{2\,\mathrm{i}\,k}\int_{-\infty}^{x} V(y)\,dy + \mathrm{O}\big(\,|\,k\,|^{-2}\,\big). \tag{6.21}$$

In particular, we can recover the potential $V(x)$ by differentiating either of these formulas and letting $k \longrightarrow \infty$:

$$V(x) = -2\lim_{k\to\infty}\left(\,\mathrm{i}\,k\,\frac{\partial}{\partial x}\big[\,e^{\,\mathrm{i}\,k\,x}\,\varphi(x,k)\,\big]\right) = -2\lim_{k\to\infty}\left(\,\mathrm{i}\,k\,\frac{\partial}{\partial x}\big[\,e^{-\,\mathrm{i}\,k\,x}\,\psi(x,k)\,\big]\right). \tag{6.22}$$

For $k \neq 0$, the Jost solutions $\psi(x,k)$ and $\overline{\psi}(x,k)$ are linearly independent, since their *Wronskian*

$$W[\psi,\overline{\psi}] = \psi\,\overline{\psi}_x - \psi_x\,\overline{\psi}$$

satisfies the asymptotic estimate

$$W[\psi(x,k),\overline{\psi}(x,k)] \sim -2\,\mathrm{i}\,k + \mathrm{o}(1), \qquad x \longrightarrow +\infty.$$

However, the Wronskian of any two solutions to the Schrödinger equation is a constant, independent of x, hence we have

$$W[\psi(x,k),\overline{\psi}(x,k)] = -2\,\mathrm{i}\,k.$$

The key to the scattering theory is to write the Jost solution at $-\infty$, i.e., $\varphi(x,k)$, in terms of the Jost solutions at $+\infty$. Now these are only all defined for real values of k. Thus, we know that for real $k \neq 0$, there exist functions $a(k), b(k)$ such that

$$\varphi(x,k) = a(k)\,\overline{\psi}(x,k) + b(k)\,\psi(x,k).$$

In fact, by the asymptotic formulas, we see that

$$a(k) = \frac{W[\varphi(x,k),\psi(x,k)]}{2\,\mathrm{i}\,k}, \qquad b(k) = -\,\frac{W[\varphi(x,k),\overline{\psi}(x,k)]}{2\,\mathrm{i}\,k}, \tag{6.23}$$

We note that the reflection and transmission coefficients are given in terms of a and b by

$$R(k) = \frac{b(k)}{a(k)}, \qquad T(k) = \frac{1}{a(k)}. \tag{6.24}$$

Moreover, the function $a(k)$, which is the key to the problem, as defined by the Wronskian relation (6.23), is analytic on the upper half plane $\mathrm{Im}\,k > 0$, and continuous down to the real axis, except possibly at $k = 0$. (On the other hand, $b(k)$ is only defined on the real axis.) Note also that, on the real axis, $a(-k) = \overline{a(k)}]$. In fact, the elementary Wronskian relation

$$W[\varphi,\psi]\,W[\overline{\varphi},\overline{\psi}] - W[\varphi,\overline{\psi}]\,W[\overline{\varphi},\psi] = W[\psi,\overline{\psi}]\,W[\varphi,\overline{\varphi}]$$

implies that, for k real,

$$|\,a(k)\,|^2 - |\,b(k)\,|^2 = 1,$$

which is equivalent to the probability relation

$$| R(k) |^2 + | T(k) |^2 = 1, \tag{6.25}$$

between the reflection and transmission coefficients. This mathematical formula encapsulates the physical statement that the collision of the particle with the quantum potential is perfectly elastic; either the particle reflects or is transmitted with no loss of energy.

Theorem 6.4. *The spectrum of the Schrödinger equation is composed of the continuous spectrum, which is the positive real axis $\lambda \geq 0$ (k real) and a finite number of simple negative eigenvalues $\lambda_1 < \lambda_2 < \ldots < \lambda_n < 0$. These correspond to the zeros of $a(k)$, which are all purely imaginary: $k_j = i \kappa_j = i \sqrt{-\lambda_j}$.*

Proof: First note that the bound states of the Schrödinger equation correspond to solutions which, for $\operatorname{Im} k > 0$, are asymptotic to $e^{i k x}$ as $x \to +\infty$, and asymptotic to $e^{-i k x}$ as $x \to -\infty$. In other words $\lambda = k^2$ will be an eigenvalue if and only if the Jost solution $\varphi(x, k)$ is a multiple of the Jost solution $\psi(x, k)$. But this is equivalent to their Wronskian vanishing, so that eigenvalues of the Schrödinger equation correspond to zeros of the function $a(k)$, or, equivalently, to the poles of the reflection coefficient $R(k)$ in the complex plane. We therefore need to analyze the behavior of the analytic function $a(k)$.

First, if $\psi = \psi(x, k)$, $\widetilde{\psi} = \psi(x, \widetilde{k})$ are any two solutions corresponding to different values of the eigenvalue, we have

$$W[\psi, \widetilde{\psi}] \, \Big|_{x=a}^{b} = \left(\psi \, \widetilde{\psi}_x - \psi_x \, \widetilde{\psi} \right) \Big|_{x=a}^{b} = \int_a^b \left(\psi \, \widetilde{\psi}_{xx} - \psi_{xx} \, \widetilde{\psi} \right) dx = \left(k^2 - \widetilde{k}^2 \right) \int_a^b \psi \, \widetilde{\psi} \, dx.$$

(In particular, choosing $k = \widetilde{k}$ proves the constancy of the Wronskian.) Now, let $\widetilde{k} = -\overline{k}$, and evaluate this identity for the Jost solutions $\varphi(x, k)$, $\varphi(x, -\overline{k}) = \overline{\varphi(x, k)}$. Then

$$W\big[\varphi(x, k), \varphi(x, -\overline{k}) \big] = \left(k^2 - \overline{k}^2 \right) \int_{-\infty}^{x} | \varphi(y, k) |^2 \, dy.$$

Now, if k is an eigenvalue, then $\varphi(x, k)$ is a multiple of $\psi(x, k)$. Therefore, the Wronskian on the left hand side of the previous equation is a multiple of the Wronskian $W\big[\psi(x, k), \psi(x, -\overline{k}) \big]$. But, as $x \longrightarrow +\infty$ this latter Wronskian goes to 0, hence we conclude that if k is an eigenvalue,

$$\left(k^2 - \overline{k}^2 \right) \int_{-\infty}^{\infty} | \varphi(y, k) |^2 \, dy = 0.$$

Since φ is not identically 0, this implies that k is purely imaginary, and so $\lambda = k^2$ is real, negative. We note that it is not possible for 0 to be an eigenvalue since the solutions in this case are never exponentially decreasing at $\pm\infty$. We conclude that all the eigenvalues of the Schrödinger equation (6.15) are strictly negative.

We next show that there are only a finite number of eigenvalues, which is equivalent to proving that the function $a(k)$ has at most a finite number of zeros for $\operatorname{Im} k > 0$. This

will follow from an estimate of its asymptotic behavior as $|k| \longrightarrow \infty$. We begin with the connection formula

$$\varphi(x, k) = a(k)\, \overline{\psi}(x, k) + b(k)\, \psi(x, k) \sim a(k)\, e^{-i k x} + b(k)\, e^{i k x}, \qquad x \longrightarrow +\infty,$$

which is valid for real $k \neq 0$. On the other hand, consider our earlier integral equation for the solution:

$$\varphi(x, k) = e^{i k x} - \frac{1}{2 i k} \int_{-\infty}^{x} \left[V(y) e^{i k(x+y)} - e^{i k(x-y)} \varphi(y, k) \right] dy$$

$$= e^{i k x} \left(\frac{1}{2 i k} \int_{-\infty}^{x} V(y)\, e^{-i k y}\, \varphi(y, k)\, dy \right)$$

$$+ e^{-i k x} \left(1 - \frac{1}{2 i k} \int_{-\infty}^{x} V(y)\, e^{i k y}\, \varphi(y, k)\, dy \right).$$

Comparing these two expressions as $x \longrightarrow \infty$, we deduce that

$$a(k) = 1 - \frac{1}{2 i k} \int_{-\infty}^{\infty} V(y)\, e^{i k y}\, \varphi(y, k)\, dy,$$

$$b(k) = \frac{1}{2 i k} \int_{-\infty}^{\infty} V(y)\, e^{-i k y}\, \varphi(y, k)\, dy.$$

By analyticity, the formula for $a(k)$ is valid for $\operatorname{Im} k \geq 0$. In particular, if we use our exponential estimate on $|e^{i k x}\, \varphi(x, k)|$ we deduce the key estimate

$$a(k) = 1 - \frac{1}{2 i k} \int_{-\infty}^{\infty} V(y)\, dy + \mathrm{O}(|k|^{-2}), \qquad \operatorname{Im} k > 0.$$

Therefore, as $|k| \longrightarrow \infty$, the function $a(k)$ is asymptotic to the constant 1. Finally, we must also eliminate the possibility that $a(k)$ has zeros accumulating at $k = 0$. Under our assumption on the potential, the integral

$$M = \int_{-\infty}^{\infty} V(y)\, \varphi(y, 0)\, dy$$

exists and is finite since the solutions for $k = 0$ are asymptotic to linear functions $\alpha x + \beta$ as $x \to \pm\infty$. Then, as $k \to 0$,

$$a(k) = 1 - \frac{M}{2 i k} - \frac{1}{2 i} \int_{-\infty}^{\infty} V(y)\, \frac{e^{i k y}\, \varphi(y, k) - \varphi(y, 0)}{k}\, dy.$$

Using some further estimates on the integrand, which, in the limit as $k \longrightarrow 0$ reduces to the derivative of $e^{i k y}\, \varphi(y, k)$ with respect to k at $k = 0$, which can be bounded by $C(1 + y^2)$ for $\operatorname{Im} k \geq 0$, we conclude that if $M = 0$, then the limit $a(0)$ exists; whereas if $M \neq 0$ then a has a simple pole at $k = 0$. In either case, the zeros of a are bounded away from $k = 0$. This completes the proof that the analytic function $a(k)$ has at most a finite number of zeros.

We prove next that the zeros of $a(k)$ are all simple. First we differentiate the Schrödinger equation with respect to k:

$$-\psi_{xxk} + V\psi_k = k^2\psi_k + 2k\psi,$$

where $\psi_k = \partial\psi/\partial k$. Multiply the Schrödinger equation by ψ_k and subtract this equation multiplied by ψ:

$$\psi\,\psi_{xxk} - \psi_k\psi_{xx} = -2k\psi^2.$$

Integrating from a to b, we have

$$W[\psi,\psi_k]\Big|_{x=a}^{b} = -2k\int_a^b \psi^2\,dx.$$

In particular, for the Jost solution $\psi(x,k)$, this implies

$$\lim_{x\to-\infty} W\big[\psi(x,k),\psi_k(x,k)\big] = 2k\int_{-\infty}^{\infty}\psi(x,k)^2\,dx.$$

Now, if $k_j = i\kappa_j$ is a zero of $a(k)$ corresponding to the eigenvalue $\lambda_j = -\kappa_j^2$, then $\varphi(x,k_j) = c_j\psi(x,k_j)$ for some constant $c_j \neq 0$. On the other hand, differentiating the identity $W[\varphi,\psi] = 2ika(k)$ with respect to k, we find

$$W[\varphi_k,\psi] + W[\varphi,\psi_k] = 2ia(k) + 2ika'(k).$$

Suppose $\lambda = k^2$ is an eigenvalue, so $a(k) = 0$. Evaluating this expression at $-\infty$, using the previous formula and noting that the first term, which is proportional to $W[\varphi_k,\varphi]$ vanishes there, yields

$$ia'(k_j) = c_j\int_{-\infty}^{\infty}\psi(x,k_j)^2\,dx \neq 0.$$

Let $\chi(x)$ be the normalized real eigenfunction corresponding to the eigenvalue $\lambda_j = -\kappa_j^2$, so $\|\chi_j\|^2 = 1$. The associated *normalization constant* d_j is defined so that

$$\chi_j(x) = d_j\psi(x,k_j).$$

Since χ_j is only defined up to sign, we can choose $d_j > 0$ without loss of generality. Then

$$d_j^2 = \frac{1}{\|\psi(x,k_j)\|^2} = -i\,\frac{c_j}{a'(k_j)}\,. \tag{6.26}$$

We have now defined all the important quantities for the scattering problem.

Definition 6.5. Given a potential $V(x)$, the *scattering data* $\mathcal{S}$ consists of

(i) The reflection coefficient $R(k)$ defined for $k \geq 0$, with $R(k) = o(1)$ as $k \to \infty$.

(ii) The eigenvalues $\lambda_1,\ldots,\lambda_n$, which are negative.

(iii) The normalization constants $d_1,\ldots,d_n$, which are positive.

The basic result is that the potential is uniquely determined by its scattering data $\mathcal{S}$. Moreover, to reconstruct the potential from $\mathcal{S}$, it suffices to solve a certain linear integral equation known as the *Gel'fand-Levitan-Marchenko* (GLM) *equation*. The key to this are the transformation operators which map solutions of the trivial Schrödinger equation with zero potential to those of the Schrödinger equation with potential V. Define the Fourier transforms

$$
\begin{aligned}
K^+(x,y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\psi(x,k) - e^{ikx} \right] e^{-iky} \, dk, \qquad y \geq x, \\
K^-(x,y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\varphi(x,k) - e^{-ikx} \right] e^{iky} \, dk, \qquad y \leq x.
\end{aligned}
\tag{6.27}
$$

The existence of these functions follows from the estimates on the Jost solutions. Moreover, the Payley-Wiener Theorem from Fourier analysis proves that $K^+(x,y) = 0$ for $y < x$, and similarly $K^-(x,y) = 0$ for $y > x$. (The precise result is in [**52**], which proves that the Fourier transform $\widehat{f}(y)$ of the function $f(k)$ vanishes for $y < x$ if and only if $f(k)$ is entire for $\operatorname{Im} k > 0$, and $\displaystyle\int_{-\infty}^{\infty} |f(a+ib)|^2 \, da = \mathrm{O}(e^{-2yb})$. In our case, $\psi(x,k) - e^{ikx}$ satisfies these conditions.) Therefore, inverting the Fourier transforms, we deduce the transformation formulas

$$
\psi(x,k) = e^{ikx} + \int_x^{\infty} K^+(x,y)\, e^{iky} \, dy, \qquad \varphi(x,k) = e^{-ikx} + \int_{-\infty}^{x} K^-(x,y)\, e^{-iky} \, dy,
$$

expressing the Jost solutions as Volterra integral operators applied to the "bare solutions" $e^{\pm ikx}$. Note that both K^+ and K^- are real, since taking complex conjugates and replacing k by $-k$ leads to the same expressions. The asymptotic expansion of these integrals shows that

$$
\psi(x,k)\, e^{-ikx} = 1 + \frac{i}{k}\, K^+(x,x) + \mathrm{O}\left(|k|^{-2} \right).
$$

Comparison with our previous expression (6.20), we deduce that

$$
V(x) = 2\frac{d}{dx} K^+(x,x).
\tag{6.28}
$$

Therefore, knowledge of the kernel K^+ allows us to reconstruct the potential V. The goal now is to derive an integral equation for K^+.

Now, since $T(k) \longrightarrow 1$ and hence $R(k) \longrightarrow 0$ sufficiently rapidly as $k \longrightarrow \infty$, we can ensure the existence of the Fourier transforms

$$
\widehat{T}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[T(k) - 1 \right] e^{-iky} \, dy, \qquad \widehat{R}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(k)\, e^{iky} \, dy.
$$

We now take our fundamental physical solution

$$
T(k)\,\varphi(x,k) = \overline{\psi}(x,k) + R(k)\,\psi(x,k),
$$

which we write in the form

$$
\begin{aligned}
\left[T(k) - 1 \right] \varphi(x,k) = {}& \left[\overline{\psi}(x,k) - e^{-ikx} \right] \\
& + R(k) \left[\psi(x,k) - e^{ikx} \right] + R(k)\, e^{ikx} - \left[\varphi(x,k) - e^{-ikx} \right],
\end{aligned}
$$

which is amenable to Fourier transform. We compute the Fourier transform of the individual terms, in order of difficulty:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} R(k)\, e^{\mathrm{i}kx} e^{\mathrm{i}ky}\, dk = \widehat{R}(x+y),$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \big[\, \varphi(x,k) - e^{-\mathrm{i}kx} \,\big] e^{\mathrm{i}ky}\, dk = K^-(x,y)$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \big[\, \overline{\psi}(x,k) - e^{-\mathrm{i}kx} \,\big] e^{\mathrm{i}ky}\, dk = \overline{\frac{1}{2\pi} \int_{-\infty}^{\infty} \big[\, \psi(x,k) - e^{\mathrm{i}kx} \,\big] e^{-\mathrm{i}ky}\, dk} = K^+(x,y),$$

since K^+ is real, and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \big[\, \psi(x,k) - e^{\mathrm{i}kx} \,\big] R(k)\, e^{\mathrm{i}ky}\, dk = \int_{-\infty}^{\infty} K^+(x,z)\widehat{R}(y+z)\, dz$$

by the convolution theorem for Fourier transforms. The final integral is a bit harder. We evaluate

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \big[\, T(k) - 1 \,\big]\varphi(x,k)\, e^{\mathrm{i}ky}\, dy$$

by contour integration. Let $\Gamma = \Gamma_R$ be the contour in the upper half plane consisting of a semi-circle C_R of radius $R \gg 0$ and the line segment from $-R$ to R. Then, by the residue theorem, since $T(k) = 1/a(k)$ has poles at the eigenvalues $k_j = \mathrm{i}\kappa_j$,

$$\frac{1}{2\pi\mathrm{i}} \oint_{\Gamma_R} \big[\, T(k) - 1 \,\big]\varphi(x,k)\, e^{\mathrm{i}ky}\, dy = \sum_{j=1}^{n} \operatorname{Res} \big[\, T(k)\varphi(x,k)\, e^{\mathrm{i}ky} \,\big]\,\big|_{k=k_j}$$

$$= \sum_{j=1}^{n} \frac{\varphi(x,k_j)\, e^{\mathrm{i}k_j y}}{a'(k_j)} = \sum_{j=1}^{n} \frac{c_j\, \psi(x,k_j)\, e^{-\kappa_j y}}{a'(k_j)} = \sum_{j=1}^{n} \mathrm{i}\, d_j^2\, \psi(x,k_j)\, e^{-\kappa_j y}.$$

We then re-express ψ in terms of the kernel K^+, to finally deduce

$$\frac{1}{2\pi\mathrm{i}} \oint_{\Gamma_R} \big[\, T(k) - 1 \,\big]\varphi(x,k)\, e^{\mathrm{i}ky}\, dy = \mathrm{i} \sum_{j=1}^{n} d_j^2 \bigg[e^{-\kappa_j(x+y)} + \int_{-\infty}^{\infty} K^+(x,z)\, e^{-\kappa_j(y+z)}\, dz \bigg].$$

We now plug all these formulas back into the original Fourier transform formula. To this end, we define the function

$$F(x) = \widehat{R}(x) + \sum_{j=1}^{n} d_j^2\, e^{-\kappa_j x} = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(k)\, e^{\mathrm{i}kx}\, dk + \sum_{j=1}^{n} d_j^2\, e^{-k_j x}.$$

Note that F is entirely determined by the scattering data. Finally recall that $K^-(x,y)$ vanishes for $y > x$. Then, our formula becomes the celebrated *Gel'fand–Levitan–Marchenko* (GLM) *equation*

$$K^+(x,y) + F(x+y) + \int_{x}^{\infty} K^+(x,z)\, F(z+y)\, dz = 0, \qquad y > x. \qquad (6.29)$$

Therefore, to find the kernel $K^+(x,y)$, one needs only solve a linear integral equation. To recover the potential, we continue K^+ to the diagonal $x = y$ and use formula (6.28). Thus, the reconstruction of the potential from the scattering data is a linear procedure! There are now some technical matters to be overcome in proving the existence of a solution to the GLM equation, and verifying that the potential satisfies the growth assumption. This will be the case provided the Fourier transform of the reflection coefficient satisfies

$$\int_a^\infty |\widehat{R}(x)| \, dx < \infty \qquad \text{for} \qquad a > -\infty,$$

and one more technical condition to get the transmission coefficient behaving properly as $k \to 0$. We refer to the references for details.

Example 6.6. *Reflectionless potentials.* The GLM equation allows us to solve the following problem: characterize those potentials $V(x)$ for which there is no quantum mechanical reflection, i.e., all particles are transmitted with probability 1.

Consider first the case of just one bound state corresponding to the eigenvalue $\lambda = -\kappa^2$ with $\kappa > 0$. According to the formula for the kernel in the GLM equation, we have

$$F(x) = d^2 e^{-\kappa x}.$$

where $d > 0$ is the normalization coefficient. The GLM equation (6.29) reduces to

$$K^+(x,y) + d^2 \, e^{-\kappa(x+y)} + \int_x^\infty K^+(x,z) \, d^2 \, e^{-\kappa(y+z)} \, dz = 0, \qquad y > x.$$

To solve this, note that we can factor

$$K^+(x,y) = \gamma(x) \, e^{-\kappa y},$$

where $\gamma(x)$ satisfies

$$0 = \gamma(x) + d^2 \, e^{-\kappa x} + \int_x^\infty \gamma(x) \, d^2 \, e^{-2\kappa z} \, dz = \gamma(x) \left(1 + \frac{d^2}{2\kappa} \, e^{-2\kappa x} \right) + d^2 \, e^{-\kappa x}.$$

Therefore,

$$\gamma(x) = -\frac{d^2 \, e^{-\kappa x}}{1 + \dfrac{d^2}{2\kappa} \, e^{-2\kappa x}}, \qquad K^+(x,y) = -\frac{d^2 \, e^{-\kappa(x+y)}}{1 + \dfrac{d^2}{2\kappa} \, e^{-2\kappa x}},$$

and so the potential is

$$\begin{aligned}
V(x) &= -2 \frac{d}{dx} K^+(x,x) \\
&= 2 \frac{d}{dx} \frac{d^2 \, e^{-2\kappa x}}{1 + \dfrac{d^2}{2\kappa} \, e^{-2\kappa x}} = \frac{-4 \, \kappa \, d^2 \, e^{-2\kappa x}}{\left(1 + \dfrac{d^2}{2k} \, e^{-2kx} \right)^2} = \frac{-4 \kappa^2}{\left(\dfrac{\sqrt{2\kappa}}{d} \, e^{kx} + \dfrac{d}{\sqrt{2\kappa}} \, e^{-\kappa x} \right)^2},
\end{aligned}$$

or, more compactly,

$$V(x) = -\kappa^2 \operatorname{sech}^2 k(x - \delta), \qquad \text{where} \qquad \delta = \frac{1}{k} \log\left(\frac{d}{\sqrt{2\kappa}}\right). \qquad (6.30)$$

Note that the normalization constant only enters into the phase d of the potential. Thus, the sech^2 potential (6.30) has no quantum mechanical reflection, and sustains a single bound state.

In the case of several eigenvalues, the solution of the GLM equation also reduces to an algebraic problem, but the computations are a bit more complicated. The kernel is now

$$F(x) = \sum_{j=1}^{n} d_j^2\, e^{-k_j x}.$$

We look for a solution of the form

$$K^+(x, y) = \sum_{j=1}^{n} w_j(x)\, e^{-k_j y}.$$

Plugging into the GLM equation (6.29), and separating out the exponentials in y, we deduce that the column vector of functions $\mathbf{w}(x) = [w_1(x), \ldots, w_n(x)]^T$ must satisfy the linear system

$$M(x)\mathbf{w}(x) = -\mathbf{e}(x),$$

where

$$\mathbf{e}(x) = [e_1(x), \ldots, e_n(x)]^T \qquad \text{has entries} \qquad e_j(x) = d_j^2\, e^{-k_j x},$$

and $M(x)$ is an $n \times n$ matrix with entries

$$M_{jk}(x) = \delta_{jk} + \int_x^\infty e^{-\kappa_k z}\, d_j^2\, e^{-\kappa_j z}\, dz = \delta_{jk} + \frac{d_j^2\, e^{-(\kappa_j + \kappa_k)x}}{\kappa_j + \kappa_k}.$$

Therefore

$$\mathbf{w}(x) = -M(x)^{-1}\mathbf{e}(x).$$

The corresponding potentials are

$$V(x) = -2\frac{d}{dx}K^+(x, x) = -2\frac{d}{dx}\sum_{j=1}^{n} w_j(x)e^{-\kappa_j x} = -2\frac{d}{dx}\sum_{j,k=1}^{n} M(x)_{jk}^{-1}\, d_k^2\, e^{-(\kappa_j + \kappa_k)x}$$

$$= -2\frac{d}{dx}\operatorname{tr}\left[M(x)^{-1}\frac{dM}{dx}\right] = -2\frac{d}{dx}\left[\frac{1}{\det M(x)}\frac{d}{dx}\det M(x)\right] = -2\frac{d^2}{dx^2}\log\det M(x).$$

A slight modification of the matrix M leads to a simpler formula. Let $T(x)$ be the matrix obtained from M by multiplying the jth row by $e^{-\kappa_j x}$ and the kth column by $e^{\kappa_k x}$, so

$$T_{jk}(x) = \delta_{jk} + \frac{d_j^2\, e^{-2\kappa_j x}}{\kappa_j + \kappa_k}. \qquad (6.31)$$

Clearly $\det T(x) = \det M(x)$, hence

$$V(x) = -2 \frac{d^2}{dx^2} \log \det T(x). \tag{6.32}$$

This gives a general formula for all the reflectionless potentials. In general, these are pretty complicated, although they can be viewed as certain nonlinear combinations of the basic sech^2 potentials, each one providing one of the eigenvalues. For certain values of the normalization constants, these can be effectively separated into n disjoint one-eigenvalue potentials.

Isopectral Flows

Let $V(t, x)$ be a one-parameter family of potentials, where we identify the parameter t with time. These are called *isospectral* if they all have identical eigenvalues. The GLM equation (6.29) characterizes (at least in principle) all such potentials. The problem of isospectral flows is to construct such families as solutions to certain differential equations. There is an efficient method due to Lax for realizing an isospectral flow, [**31**].

Theorem 6.7. *Let*

$$L(t) = \partial_x^2 + V(t, x) \tag{6.33}$$

denote a one-parameter family of Schrödinger operator depending on the time-varying potential V. Suppose B is a skew adjoint operator such that the commutator $[B, L]$ is a multiplication operator. Then the flow determined by the Lax representation

$$L_t = [B, L] \tag{6.34}$$

is isospectral.

Note that the Lax representation will reduce to a (typically nonlinear) evolution equation

$$\frac{\partial V}{\partial t} = K[V] \tag{6.35}$$

for the potential $V(t, x)$, where K is a certain operator arising from the form of the commutator: $K[V] = [B, L]$.

Proof: Let $L(t)$ be a solution to the Lax equation (6.34). Then these operators are unitarily equivalent, meaning that there is a one-parameter family of unitary operators $U(t)$ such that

$$L(t) = U(t) L(0) U(t)^{-1}. \tag{6.36}$$

Indeed, by Theorem 4.1, the skew adjoint operator B generates a one-parameter unitary group of operators

$$U_t = B U, \qquad U(0) = \mathrm{I}.$$

Defining $L(t)$ by (6.36), we see that

$$L_t = B U L(0) U^{-1} - U L(0) U^{-1} B = B L - L B,$$

so that $L(t)$ satisfies the Lax equation **e.qLax** Suppose λ is an eigenvalue for $L(0)$ corresponding to the eigenfunction $\psi_0(x)$. Then λ is also an eigenvalue for $L(t)$ corresponding to the eigenfunction $\psi(t, x) = U(t)\,\psi_0(x)$, because

$$L(t)\,\psi(t, x) = (U(t)\,L(0)\,U(t)^{-1})\,U(t)\,\psi_0(x) = U(t)\,L(0)\,\psi_0(x) = \lambda\,U(t)\,\psi_0(x) = \lambda\,\psi(t, x).$$

This completes the proof. *Q.E.D.*

A simple example is provided by the operator $B = \partial_x$. Note that

$$[B, L] = V_x$$

is the operator of multiplication by the derivative of the potential. The corresponding isospectral flow (6.35) coincides with the unidirectional wave equation

$$V_t = V_x,$$

so we deduce the easy result that any translation of a potential has the same eigenvalues. The next example is the third order differential operator

$$B = \partial_x^3 + \tfrac{3}{2} V\,\partial_x + \tfrac{3}{4} V_x. \tag{6.37}$$

Note that B is skew-adjoint. A short computation shows that

$$[B, L] = \tfrac{1}{4} V_{xxx} + \tfrac{3}{2} V\,V_x$$

is also a multiplication operator. The corresponding isospectral flow (6.35) is the celebrated *Korteweg-deVries equation*

$$V_t = \tfrac{1}{4} V_{xxx} + \tfrac{3}{2} V\,V_x,$$

which models long surface waves on shallow water and a broad range of other physical phenomena, [**57**]. In fact, there are operators B of all odd orders, which in turn give rise to higher order isospectral flows, [**24**]. All of these flows are, in fact, completely integrable, possessing an infinite number of symmetries and conservation laws, and can thus can be viewed as infinite-dimensional counterparts of the classical completely integrable Hamiltonian systems.

Moreover, these isospectral evolution equations can all be solved by the method of inverse scattering. The eigenvalues of the corresponding Schrödinger operator are constant, independent of t. The rest of the scattering data, while not constant, has a rather elementary time evolution. Note that the (generalized) eigenfunctions will evolve in time according to $\psi_t = B\,\psi$, where B is the Lax operator. Taking the Korteweg-deVries example (6.37), we have

$$\psi_t = \psi_{xxx} + \tfrac{3}{2} V\,\psi_x + \tfrac{3}{4} V_x\,\psi.$$

Now suppose that $\psi(x)$ is the physical solution

$$\psi(x) \;\sim\; \begin{cases} e^{-ikx} + R(k)\,e^{ikx}, & x \longrightarrow +\infty, \\ T(k)\,e^{-ikx}, & x \longrightarrow -\infty, \end{cases}$$

as before. For $|x|$ large, $V(x)$ is negligible, and hence ψ evolves according to the linear equation $\psi_t = \psi_{xxx}$. If the initial data $\psi(0, x)$ is physical, then $\psi(t, x)$ will be a multiple of this solution

$$\psi(t, x) \ \sim \ \begin{cases} e^{\mathrm{i}\,k^3 t}\, e^{-\mathrm{i}\,k x} + R(k)\, e^{-\mathrm{i}\,k^3 t}\, e^{\mathrm{i}\,k x}, & x \longrightarrow +\infty, \\ T(k)\, e^{\mathrm{i}\,k^3 t}\, e^{-\mathrm{i}\,k x}, & x \longrightarrow -\infty, \end{cases}$$

which is $e^{\mathrm{i}\,k^3 t}$ times the physical solution for $V(t, x)$. Therefore, we conclude that the transmission coefficient $T(t, k)$ corresponding to $V(t, x)$ is constant in time, while the reflection coefficient has the elementary time evolution

$$R(t, k) = e^{-2\,\mathrm{i}\,k^3 t} R(0, k).$$

Similarly, the normalization constants were defined as $d_j = \| \, \psi(x, k_j) \, \|^{-1}$, hence

$$d_j(t) = e^{2\,k_j^3 t} d_j(0).$$

Therefore, to solve the Cauchy problem for the Korteweg-deVries equation:

$\quad (i)$ Compute the scattering data $\mathcal{S}(0)$ for the initial potential $V(0, x)$.
$\quad (ii)$ Determine the time evolution for the scattering data $\mathcal{S}(t)$ as above.
$\quad (iii)$ Solve the GLM equation to determine $V(t, x)$.

For example, the n-soliton solutions correspond to the case when the reflection coefficient is identically 0, and so these can be explicitly found as above. In particular, the one-soliton solution corresponds to the reflectionless potentials with a single bound state, and hence have the sech^2 profile (6.30) moving to the left as a traveling wave unchanged in form and with constant speed. Moreover, the larger the amplitude of a soliton, the faster it moves. For $n \geq 2$, it can be shown that, as $t \to +\infty$, the n-soliton solutions break up into a distinct collection of one-soliton traveling wave solutions, ordered by size, each soliton corresponding to one of the eigenvalues. As $t \to -\infty$, a similar break up occurs, but the solitons are in reverse order. Also, there is a phase shift of the individual solitons reflecting the effect of the nonlinear collision. See [1, 21] for further developments.

7. Symmetry in Quantum Mechanics.

References: [9], [36], [37], [39], [42], [49], [53], [54], [55].

Why is Symmetry Important?

For a classical mechanical system, a *symmetry group* is a group of transformations which leaves the Hamiltonian function invariant[†]. Thus, for example, the group of rotations in space leaves the Kepler Hamiltonian invariant, and the group of simultaneous

---

[†] More accurately, this defines a *canonical symmetry group*. The symmetry condition requires that the group map solutions to solutions, but there may also be non-canonical symmetries not of the form discussed here. See [42] for full details.

translations leaves the Hamiltonian energy for the n-body problem invariant. The most important symmetry groups are continuous groups of canonical transformations. Consider a one-parameter group

$$(p, q) \longmapsto (p(\varepsilon), q(\varepsilon))$$

generated by the Hamiltonian function $F(p, q)$, meaning that as functions of the group parameter ε, the transformations are solutions to Hamilton's equations (2.1) with Hamiltonian F and ε playing the role of t. The condition that the resulting transformations form a symmetry group (for H time-independent) is that their Poisson bracket vanish:

$$\{F, H\} = 0, \tag{7.1}$$

which is equivalent to the *conservation law* that $F(p, q) =$ constant. For example, translation invariance of the Hamiltonian is equivalent to conservation of linear momentum, and rotational invariance of the Hamiltonian is equivalent to conservation of angular momentum. Invariance under time translations is equivalent to conservation of energy, where H does not depend explicitly on t. Thus, there is an intimate relationship between symmetry and conservation laws. This is the essence of Noether's Theorem.

For the corresponding quantum mechanical system, the classical Hamiltonian H and the associated conserved function F generating a symmetry group are replaced by self-adjoint operators $\mathcal{H}$ and $\mathcal{F}$ on an appropriate Hilbert space $\mathcal{E}$. The Poisson bracket symmetry condition (7.1) is replaced by the commutation condition

$$[\mathcal{H}, \mathcal{F}] = 0. \tag{7.2}$$

(For simplicity, we ignore potential technical difficulties with domains of definition of these operators.) As a first consequence of this condition, we see that the physical quantity represented by the operator $\mathcal{F}$ is conserved in the quantum mechanical equation; indeed, if $\psi(0, x)$ is an eigenfunction for $\mathcal{F}$ with eigenvalue μ, and $\psi(t, x)$ is the solution of the Schrödinger equation for $\mathcal{H}$, then $\psi(t, x)$ remains an eigenfunction for $\mathcal{F}$ with eigenvalue μ. We also discover that the operator $\mathcal{F}$ maps the space of solutions to the stationary Schrödinger equation

$$\mathcal{E}_\lambda = \{\, \psi \mid \mathcal{H}\psi = \lambda\psi \,\}$$

to itself (where defined): $\mathcal{F}(\mathcal{E}_\lambda) \subset \mathcal{E}_\lambda$. This means that we have a "representation" of the group on this space, and can thus apply the powerful methods of group representation theory to study the solutions to the Schrödinger equation. Indeed, representation theory enables us to deduce consequences of the symmetry conditions which do not depend on the particular quantum mechanical realization of the physical system, but only on the pure symmetry properties of the Hamiltonian. This is similar to our earlier treatment of the harmonic oscillator, where most of the the quantum mechanics was derived abstractly without knowledge of the particular realization in physical coordinates. This will be particularly important when we generalize the Schrödinger equation to wave functions with more than one component, and, indeed, various physically observed phenomena will necessitate such a treatment. In this case, the underlying symmetry and representation theory will serve as very important guides to the construction of physically meaningful models.

Initially, the two most important groups are the rotation group (and its covering group) and the discrete group of permutation of identical particles. However, many other symmetries arise naturally, e.g., crystallographic groups, the Lorentz group in the relativistic regime, the Heisenberg group, which appears in the harmonic oscillator, etc. Whereas continuous symmetry groups in quantum mechanics always have classical analogues, based on the associated infinitesimal generators, discrete groups and certain global aspects of continuous groups often do not.

Group Representations

Let G be a group. By a *representation* of G we mean a (continuous) group homomorphism

$$\rho : G \ \longrightarrow \ \mathcal{L}(V)$$

to the space of (bounded) linear transformations of a (complex, normed) vector space V, satisfying

$$\rho(g \cdot h) = \rho(g) \cdot \rho(h), \qquad \rho(g^{-1}) = \rho(g)^{-1}, \qquad g, h \in G. \tag{7.3}$$

If V is finite-dimensional, and hence we can identify $V \simeq \mathbb{C}^n$ where $n = \dim V$, the representation ρ maps G to the group $\mathrm{GL}(n)$ of complex $n \times n$ matrices in such a way that group multiplication gets mapped to matrix multiplication. Two representations ρ on V and σ on W are called *equivalent* if there is a (bounded) invertible linear transformation $A : V \longrightarrow W$ such that $\rho(g) = A^{-1}\sigma(g)A$ for all $g \in G$. In the finite-dimensional case, this means V and W have the same dimension, and A just amounts to a change of basis.

By the *direct sum* of two representations ρ_1, ρ_2, we mean the representation

$$\rho = \rho_1 \oplus \rho_2 : G \ \longrightarrow \ \mathcal{L}(V_1 \oplus V_2), \qquad \rho(g) = \rho_1(g) \oplus \rho_2(g), \tag{7.4}$$

on the direct sum vector space. In finite dimensions, this means that the matrix representing $\rho(g)$ assumes block diagonal form with respect to the appropriate bases. A subspace $W \subset V$ is called *invariant* if $\rho(g)W \subseteq W$ for all $g \in G$. Note that the restriction of ρ to any invariant subspace is also a representation of G. A representation ρ is called *irreducible* if the only closed invariant subspaces are the trivial ones, namely $\{0\}$ and V itself. A representation is called *completely reducible* if it decomposes as a direct sum of irreducible representations. (In the infinite-dimensional situation, we mean Hilbert direct sums.)

Theorem 7.1. *If G is a finite group, or a compact Lie group, then every irreducible representation is finite-dimensional. Moreover, every (continuous) representation is completely reducible.*

The compact case is a consequence of the celebrated Peter–Weyl Theorem.

A representation ρ on an inner product space V is called *unitary* if it maps

$$\rho : G \ \longrightarrow \ \mathcal{U}(V)$$

to the space of unitary linear transformations on V, meaning that

$$\langle\, \rho(g)\psi \,, \rho(g)\psi \,\rangle = \langle\, \psi \,, \psi \,\rangle \qquad \text{for all} \qquad g \in G, \quad \psi \in V. \tag{7.5}$$

It is not difficult to show that every unitary representation is completely reducible, as a consequence of the following result.

Proposition 7.2. *Let ρ be a unitary representation of G. If W is an invariant subspace, then its orthogonal complement $W^\perp$ is also invariant, and hence ρ decomposes as the direct sum of two smaller representations: $\rho = \rho \mid W \oplus \rho \mid W^\perp$.*

The complete reducibility of all representations of finite and compact Lie groups follows from the fact that all their representations are equivalent to unitary representations. This is not true for non-compact groups; see below.

Theorem 7.3. *If G is a finite group, or a compact Lie group, then every (continuous) representation is equivalent to a unitary representation.*

Proof: In the finite case, we define the new inner product

$$(\varphi; \psi) = \frac{1}{\#G} \sum_{g \in G} \langle \rho(g)\varphi , \rho(g)\psi \rangle, \tag{7.6}$$

on V, where $\#G$ denotes the number of elements of G. It is straightforward to check that (7.6) defines an inner product, and each $\rho(h)$ for $h \in G$ then is a unitary map. For G a compact Lie group, we replace the summation in (7.6) by integration with respect to the invariant (Haar) measure.

To deduce equivalence, we let A be the map taking the orthonormal basis of V in the new inner product to that in the old. In the infinite-dimensional case, we also need to show that A is continuous, which is the same as the two inner products being equivalent. *Q.E.D.*

Example 7.4. Let $G = \mathrm{SO}(2)$ be the group of rotations in the plane $\mathbb{R}^2$. Since every rotation in the plane is uniquely prescribed by its angle, $\mathrm{SO}(2)$ is topologically a circle, S^1, parametrized by the angular coordinate θ. Since $\mathrm{SO}(2)$ is abelian, every irreducible representation is one-dimensional. (This is because commuting unitary operators can be simultaneously diagonalized.) The representations are given by

$$\rho_n(\theta) = e^{\,\mathrm{i}\,n\,\theta}, \qquad \text{where} \qquad n \in \mathbb{Z}. \tag{7.7}$$

Any other (complex) representation can be decomposed into a direct sum of these irreducible representations. For instance, the standard representation in terms of rotation matrices

$$\rho(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

can be rewritten in terms of the complex eigenvectors $e_1 \pm \mathrm{i}\,e_2$ as the direct sum of the irreducible representations corresponding to $n = 1$ and $n = -1$:

$$\rho(\theta) = \rho_1(\theta) \oplus \rho_{-1}(\theta) = \begin{pmatrix} e^{\,\mathrm{i}\,n\,\theta} & 0 \\ 0 & e^{-\,\mathrm{i}\,n\,\theta} \end{pmatrix}.$$

Similarly, the representation of $\mathrm{SO}(2)$ on the space of quadratic polynomials $q(x, y)$ is given by

$$\rho^{(2)}(\theta)\,q(x,y) = q(\rho(\theta)^{-1}(x,y)) = q(x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta).$$

It takes the matrix form

$$\rho^{(2)}(\theta) = \begin{pmatrix} \cos^2\theta & -\cos\theta\,\sin\theta & \sin^2\theta \\ 2\cos\theta\,\sin\theta & \cos^2\theta - \sin^2\theta & -2\cos\theta\,\sin\theta \\ \sin^2\theta & \cos\theta\,\sin\theta & \cos^2\theta \end{pmatrix}$$

relative to the standard monomial basis consisting of x^2, xy, y^2. It can be decomposed into the following direct sum of irreducible representations (check!):

$$\rho^{(2)} = \rho_2 \oplus \rho_0 \oplus \rho_{-2}.$$

From now on, we will restrict our attention to finite-dimensional representations so as to avoid problems with unbounded operators, etc. Also, unless otherwise stated, all representations will be on complex vector spaces, even though the Lie group may be real. The following technical result is absolutely crucial.

Schur's Lemma 7.5. *Let ρ be a complex unitary representation of G on the finite-dimensional vector space V. Then ρ is irreducible if and only if the only linear transformations $A: V \longrightarrow V$ satisfying*

$$\rho(g)\,A = A\,\rho(g) \qquad \text{for all} \qquad g \in G, \tag{7.8}$$

are multiples of the identity: $A = \lambda\,\mathrm{I}$. Furthermore, if ρ and $\widetilde{\rho}$ are two irreducible representations of G on spaces $V, \widetilde{V}$ respectively, and $B: V \longrightarrow \widetilde{V}$ is a linear map such that

$$\widetilde{\rho}(g)\,B = B\,\rho(g) \qquad \text{for all} \qquad g \in G, \tag{7.9}$$

then either B is invertible, so that ρ and $\widetilde{\rho}$ are equivalent representations, or $B = 0$.

Proof: To prove the first statement, let λ be an eigenvalue of A, and let $\{0\} \neq W \subset V$ be the associated eigenspace. It is easy to see that the condition above implies that W is invariant under ρ. Therefore, if ρ is irreducible, $W = V$, and $A = \lambda\,\mathrm{I}$. (Note that this part does not require unitarity of the representation ρ.) Conversely, if ρ is reducible, there is a nontrivial invariant subspace $\{0\} \neq W \subsetneq V$, whose orthogonal complement is also invariant. Setting A to be orthogonal projection of V onto W, it is easy to check that condition (7.8) holds.

To prove the second statement, it is easy to see that $\ker B$ is an invariant subspace of V under ρ. Thus, since ρ is irreducible, either $\ker B = V$, in which case $B = 0$, or $\ker B = \{0\}$. In the latter case, $\mathrm{rng}\,B$ is also seen to be an invariant subspace of $\widetilde{V}$, hence, by irreducibility, $\mathrm{rng}\,B = \widetilde{V}$, and B is invertible. *Q.E.D.*

Lie Algebra Representations

Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. By a *Lie algebra representation*, we mean a linear map

$$\rho : \mathfrak{g} \longrightarrow \mathfrak{gl}(V),$$

to the Lie algebra $\mathfrak{gl}(V)$ consisting of all linear operators on the vector space V (i.e., the space of all $n \times n$ matrices if V is n-dimensional) which respects the Lie bracket. Explicitly,

$$\rho[A, B] = [\rho(A), \rho(B)] = \rho(A)\,\rho(B) - \rho(B)\,\rho(A), \qquad (7.10)$$

where the right hand bracket is the matrix commutator: $[A, B] = AB - BA$. If G is a Lie group, and ρ a representation, then ρ induces a representation of the associated Lie algebra $\mathfrak{g}$, also, for simplicity, denoted ρ. Explicitly, if $A \in \mathfrak{g}$ generates the one-parameter subgroup $\exp(tA) \subset G$, then

$$\rho(A) = \left. \frac{d}{dt}\,\rho(\exp tA) \right|_{t=0}.$$

This follows easily from the standard commutator formula for the Lie bracket:

$$[A, B] = \left. \frac{d}{dt}\,\exp\!\left(-\sqrt{t}\,B\right)\,\exp\!\left(-\sqrt{t}\,A\right)\,\exp\!\left(\sqrt{t}\,B\right)\,\exp\!\left(\sqrt{t}\,A\right) \right|_{t=0}.$$

which also works for matrices. Conversely, if ρ is any representation of the Lie algebra $\mathfrak{g}$, then exponentiation induces a local representation of the corresponding Lie group G:

$$\rho(\exp A) = \exp \rho(A),$$

using the fact that the exponential map $\exp \colon \mathfrak{g} \to G$ is one-to-one in a neighborhood of the identity, which defines the Lie group representation in such a neighborhood. If G is connected, then every element can be written as a finite product of such exponentials, and hence each representation of a connected Lie group is uniquely prescribed by its associated Lie algebra representation. However, a Lie algebra representation may not extend to a Lie group representation!

For example, the one-dimensional representation of the one-dimensional Lie algebra $\mathfrak{so}(2) \simeq \mathbb{R}$, with generator J, given by

$$\rho(J) = 1,$$

does not extend to a representation of the group $\mathrm{SO}(2)$. Indeed,

$$\rho(e^{tJ}) = e^{t\rho(J)} = e^{t}, \qquad \text{but} \qquad e^{2\pi J} = \mathrm{I}.$$

This representation does extend to the simply connected covering group of $\mathrm{SO}(2)$, which is just the (non-compact) additive group of real numbers $G = \mathbb{R}$:

$$\rho(t) = e^{t}.$$

It is an example of a non-unitary representation of a non-compact group which cannot be made unitary by any choice of metric: there is no inner product on $\mathbb{R}^1$ such that

$$\| e^{t}\psi \| = \| \psi \| \qquad \text{for all } \psi \in \mathbb{R}^1 \text{ and all } t \in \mathbb{R}.$$

The obstructions to extending a local representation induced from a Lie algebra representation to the entire Lie group are topological in nature. If we were to work with

just simply connected Lie groups, then every Lie algebra representation would extend to a Lie group representation and there would be a one-to-one correspondence between the two. However, many of the important groups, e.g., the group of rotations, are not simply connected, and, moreover, by going to the simply connected covering group, we may well lose compactness.

A Lie group representation ρ is irreducible if and only if the corresponding Lie algebra representation is irreducible. The Lie group representation ρ is unitary if and only if the corresponding Lie algebra representation is skew-adjoint, i.e.,

$$\rho : \mathfrak{g} \longrightarrow \mathfrak{u}(V),$$

maps $\mathfrak{g}$ to the space $\mathfrak{u}(V)$ of skew-adjoint operators on the inner product space V, using the fact, cf. Theorem 4.1, that a skew-adjoint operator generates a one-parameter unitary group. In the finite-dimensional case, where $V \simeq \mathbb{C}^n$, by using an orthonormal basis so that the inner product is in canonical form, $\mathfrak{u}(V) \simeq \mathfrak{u}(n)$ can be identified as the Lie algebra of skew-adjoint, complex matrices under the commutator bracket.

Representations of the Orthogonal Group

The first important case of a non-commutative compact group is the *special orthogonal group* of rotations in three-dimensional space:

$$\mathrm{SO}(3) = \left\{ A \mid A^T A = \mathrm{I}, \quad \det A = 1 \right\}. \tag{7.11}$$

Its Lie algebra $\mathfrak{so}(3)$ can be identified with the space of skew-symmetric matrices, and is spanned by the matrices

$$L_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad L_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \qquad L_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7.12}$$

generating the one-parameter subgroups of rotations around the z, y, x axes respectively. They satisfy the commutation relations

$$[L_x, L_y] = L_z, \qquad [L_z, L_x] = L_y, \qquad [L_y, L_z] = L_x. \tag{7.13}$$

Note that these are the same as the Poisson bracket relations among the three components of angular momentum[†]

$$\mathbf{j} = \mathbf{q} \wedge \mathbf{p}.$$

The corresponding quantum operators are

$$\mathcal{J} = \left(\mathcal{J}_x, \mathcal{J}_y, \mathcal{J}_z \right) = -\,\mathrm{i}\,\hbar\,\mathbf{q} \wedge \nabla, \tag{7.14}$$

which satisfy the commutation relations

$$[\mathcal{J}_x, \mathcal{J}_y] = \mathrm{i}\,\hbar\,\mathcal{J}_z, \qquad [\mathcal{J}_z, \mathcal{J}_x] = \mathrm{i}\,\hbar\,\mathcal{J}_y, \qquad [\mathcal{J}_y, \mathcal{J}_z] = \mathrm{i}\,\hbar\,\mathcal{J}_x. \tag{7.15}$$

[†] As above, we use $\wedge$ to denote the vector cross product.

Note that if we rescale by $i\hbar$, setting $\mathcal{J}_\nu = i\hbar\mathcal{L}_\nu$, we recover the commutation relations of the Lie algebra $\mathfrak{so}(3)$, and hence the quantum angular momentum operators give a representation of the Lie algebra $\mathfrak{so}(3)$ on (a subspace of) the Hilbert space of state vectors. (Note that the angular momentum operators $\mathcal{J}_\nu$ are self-adjoint, whereas the generators of $\mathfrak{so}(3)$ are skew-symmetric matrices, so the factor of i is essential.)

Because the angular momentum operators do not commute, according to the Uncertainty Principle, we cannot simultaneously measure all three components (except in the case when all three vanish). Indeed, according to the general Uncertainty Relation (4.27),

$$\Delta\mathcal{J}_\alpha\,\Delta\mathcal{J}_\beta \geq \tfrac{1}{2}\hbar\langle\mathcal{J}_\gamma\rangle, \qquad \alpha,\beta,\gamma \text{ all different.}$$

Thus, if we know one component, say $\mathcal{J}_x$ exactly, so $\Delta\mathcal{J}_x = 0$, then measurements of the other two components are completely random, except in the special case when they all vanish. This poses an interesting conceptual conundrum: the angular momentum "vector" can never have a well-defined direction, in contrast to the position and momentum vectors. The only time it can be well-defined is when it is zero.

The square of angular momentum is the operator

$$\mathcal{J}^2 = \mathcal{J}_x^2 + \mathcal{J}_y^2 + \mathcal{J}_z^2. \tag{7.16}$$

It commutes with all three angular momentum operators. Therefore, one can measure the square of angular momentum and one component of angular momentum exactly, but not the other two components! Similarly, for a system composed of a number of particles, the total angular momentum and its square obey the same commutation relations as the individual pieces.

It is customary in quantum mechanics to ignore the particular realization and define an *angular momentum operator* to be a vector of self-adjoint operators

$$\mathbf{J} = (J_x, J_y, J_z),$$

which obey the commutation relations

$$[J_x, J_y] = i\,J_z, \qquad [J_z, J_x] = i\,J_y, \qquad [J_y, J_z] = i\,J_x, \tag{7.17}$$

where we suppress the Planck factor of $\hbar$ by either choosing units in which it is 1 or by rescaling the operators. This includes the case of a single particle, as well as systems of particles, where $\mathbf{J}$ represents (up to a factor of $\hbar$) the total angular momentum of the system. If we set $\mathbf{J} = i\,\mathbf{L}$, then the components of $\mathbf{L}$ are skew adjoint operators, and comprise a representation of the Lie algebra $\mathfrak{so}(3)$ satisfying the commutation relations (7.12). Therefore, according to our general results, the underlying Hilbert space decomposes into a direct sum of irreducible finite-dimensional subspaces. If our Schrödinger equation is rotationally invariant, and so the Hamiltonian operator commutes with the angular momentum operator $\mathbf{J}$, then the same is true for the solution space for each value of the energy. Thus, in order to understand the angular momentum of the system, we need to understand the irreducible representations of $\mathfrak{so}(3)$.

Their construction is facilitated by introducing the following alternative basis of the (complexification of) $\mathfrak{so}(3)$:

$$J_+ = J_x + i\,J_y, \qquad J_- = J_x - i\,J_y, \qquad J_z. \tag{7.18}$$

Observe that these operators satisfy

$$J_- = (J_+)^*, \tag{7.19}$$

along with the commutation relations

$$[J_+, J_-] = 2\,J_z, \qquad [J_z, J_+] = J_+, \qquad [J_z, J_-] = -\,J_-. \tag{7.20}$$

The operators J_+, J_- are called, respectively, *raising* and *lowering operators*, for reasons that will soon become clear. (This will also be reminiscent of our treatment of the harmonic oscillator.) Further,

$$J^2 = J_x^2 + J_y^2 + J_z^2 = \tfrac{1}{2}\,(J_+J_- + J_-J_+) + J_z^2, \tag{7.21}$$

Thus, $(7.20, 21)$ imply

$$J_-J_+ = J^2 - J_z(J_z + 1), \qquad J_+J_- = J^2 - J_z(J_z - 1). \tag{7.22}$$

Now, suppose we are looking for a (necessarily finite-dimensional) irreducible representation of these operators. (For simplicity, we suppress ρ in the formulas.) Since each J_ν^2 is a positive definite self-adjoint operator, their sum J^2 is also positive definite, and hence has only non-negative eigenvalues. For later convenience, these eigenvalues will be denoted as $\lambda = j(j+1)$, where j is known as the *azimuthal quantum number*. Since J^2 also commutes with J_z, we can find a simultaneous eigenvector $\psi_{j,l}$, without loss of generality assumed to be normalized, $\|\psi_{j,l}\| = 1$, associated with the eigenvalues $j(j+1)$ and l, respectively:

$$J^2\psi_{j,l} = j(j+1)\,\psi_{j,l}, \qquad J_z\psi_{j,l} = l\,\psi_{j,l}, \tag{7.23}$$

where l is called the *magnetic quantum number*. Set

$$\widehat{\psi}_{j,l+1} = J_+\psi_{j,l}. \tag{7.24}$$

Then, as the notation indicates,

$$J^2\,\widehat{\psi}_{j,l+1} = j(j+1)\,\widehat{\psi}_{j,l+1}, \qquad J_z\widehat{\psi}_{j,l+1} = (l+1)\,\widehat{\psi}_{j,l+1}.$$

Thus $\widehat{\psi}_{j,l+1}$, if non-zero, is also a simultaneous eigenvector with corresponding eigenvalues $j(j+1)$ and $l+1$. A similar statement holds for

$$\widehat{\psi}_{j,l-1} = J_-\psi_{j,l}, \tag{7.25}$$

which has corresponding eigenvalues $j(j+1)$ and $l-1$. The net effect is that J_+ raises the magnetic quantum number l by 1, while J_- lowers it by 1.

Now consider

$$\begin{aligned}
J_-\widehat{\psi}_{j,l+1} = J_-J_+\psi_{j,l} &= J^2\psi_{j,l} - J_z(J_z + 1)\,\psi_{j,l} \\
&= \big[\,j(j+1) - l(l+1)\,\big]\,\psi_{j,l} = (j-l)(j+l+1)\,\psi_{j,l}.
\end{aligned}$$

By a similar computation,

$$J_+\widehat{\psi}_{j,l-1} = (j+l)(j-l+1)\,\psi_{j,l}.$$

Therefore, using (7.19),

$$\| \widehat{\psi}_{j,l+1} \|^2 = \langle J_+\psi_{j,l}, J_+\psi_{j,l} \rangle = \langle \psi_{j,l}, J_-J_+\psi_{j,l} \rangle = (j-l)(j+l+1)\| \psi_{j,l} \|^2$$
$$= (j-l)(j+l+1),$$
$$\| \widehat{\psi}_{j,l-1} \|^2 = \langle J_-\psi_{j,l}, J_-\psi_{j,l} \rangle = \langle \psi_{j,l}, J_+J_-\psi_{j,l} \rangle = (j+l)(j-l+1)\| \psi_{j,l} \|^2$$
$$= (j+l)(j-l+1).$$

Since $\| \widehat{\psi}_{j,l+1} \|$, $\| \widehat{\psi}_{j,l-1} \| \geq 0$, we deduce

$$(j-l)(j+l+1) \geq 0, \qquad (j+l)(j-l+1) \geq 0,$$

which imply

$$-j \leq l \leq j.$$

Moreover $\widehat{\psi}_{j,l+1} = J_+\psi_{j,l} = 0$ if and only if $l = j$, while $\widehat{\psi}_{j,l-1} = J_-\psi_{j,l} = 0$ if and only if $l = -j$. Thus, when $|l| < j$, we can normalize the two new eigenvectors by setting

$$\psi_{j,l+1} = \frac{\widehat{\psi}_{j,l+1}}{\sqrt{(j-l)(j+l+1)}}, \qquad \psi_{j,l-1} = \frac{\widehat{\psi}_{j,l-1}}{\sqrt{(j+l)(j-l+1)}}.$$

Now, if we start with $\psi_{j,l}$ and apply J_+ successively, we produce a sequence of eigenvectors $0 \neq \psi_{j,l+k}$, $k = 0, 1, 2, \ldots$ unless $l+k = j$. Therefore, to maintain finite-dimensionality, we must require $j-l$ to be an integer. Similarly, if we start applying J_-, we infer that $j+l$ is also an integer. Adding these two integers, we deduce that $2j$ must be an integer, so j must be a non-negative integer: $0, 1, 2, 3, \ldots$, or half-integer: $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$. In this fashion, we have constructed all the irreducible representations of $\mathfrak{so}(3)$.

Theorem 7.6. *Every irreducible representation of the Lie algebra $\mathfrak{so}(3)$ is characterized by the azimuthal quantum number j which can assume any non-negative integral or half-integral value:*

$$j = 0, \ \tfrac{1}{2}, \ 1, \ \tfrac{3}{2}, \ 2, \ \tfrac{5}{2}, \ \ldots \ .$$

The representation ρ_j corresponding to a given value of j has dimension $2j+1$, and is characterized by the fact that the squared angular momentum operator J^2 has eigenvalue $j(j+1)$. There is an orthonormal basis of corresponding representation space $V_j \simeq \mathbb{C}^{2j+1}$ provided by the eigenvectors $\psi_{j,l}$ for $l = -j, -j+1, \ldots, j-1, j$, of the operator J_z, satisfying

$$\begin{aligned} J^2\psi_{j,l} &= j(j+1)\,\psi_{j,l}, & J_+\psi_{j,l} &= \sqrt{j(j+1)-l(l+1)}\ \psi_{j,l+1}, \\ J_z\psi_{j,l} &= l\,\psi_{j,l}, & J_-\psi_{j,l} &= \sqrt{j(j+1)-l(l-1)}\ \psi_{j,l-1}. \end{aligned} \qquad (7.26)$$

For example, $j = 0$ is the trivial one-dimensional representation in which $\mathbf{J} = 0$. The case $j = 1$ is equivalent to the standard three-dimensional representation of the orthogonal group.

Consider the standard realization of the generators of $\mathfrak{so}(3)$ as the quantum operators of angular momentum,

$$\mathbf{J} = -\,\mathrm{i}\,\mathbf{q} \wedge \nabla, \tag{7.27}$$

acting on the Hilbert space $\mathcal{E} = \mathrm{L}^2(\mathbb{R}^3)$ of square integrable functions of $\mathbf{q} = (x, y, z) \in \mathbb{R}^3$. If we go to spherical coordinates (r, φ, θ), as in (2.25), then the constituent irreducible representations are spanned by joint eigenfunctions of the operators

$$J_z = -\,\mathrm{i}\,\frac{\partial}{\partial \theta}, \qquad J^2 = -\left(\frac{1}{\sin \varphi}\,\frac{\partial}{\partial \varphi}\,\sin \varphi\,\frac{\partial}{\partial \varphi} + \frac{1}{\sin^2 \varphi}\,\frac{\partial^2}{\partial^2 \theta} \right). \tag{7.28}$$

The operator J^2 is known as the *spherical Laplacian*, since it ca be identified as the *Laplace–Beltrami operator* on the unit sphere $S^2 \subset \mathbb{R}^3$ with induced constant curvature Riemannian metric.

We now seek the eigenfunction $\psi_{j,l}(\varphi, \theta)$ satisfying (7.23). (When looking for these eigenfunctions, we can safely ignore the radial coordinate.) Now, J_z is just differentiation in the θ direction, hence

$$\psi_{j,l}(\varphi, \theta) = \chi_{j,l}(\varphi)\, e^{\,\mathrm{i}\, l\theta}.$$

Since θ is periodic of period 2π, the magnetic quantum number m must be an integer, which implies that the azimuthal quantum number j must be an integer too. Therefore, for the standard angular momentum operators (7.27), there are no half-integral representations! The functions $\chi_{j,l}(\varphi)$ must satisfy the ordinary differential equation

$$-\frac{1}{\sin \varphi}\,\frac{d}{d\varphi}\,\sin \varphi\,\frac{d}{d\varphi}\,\chi_{j,l} + \frac{l^2}{\sin^2 \varphi}\,\chi_{j,l} = j(j+1)\,\chi_{j,l}.$$

If we set $u = \cos \varphi$, then this becomes the *Legendre equation*

$$\left[(1 - u^2)\frac{d^2}{du^2} - 2u\frac{d}{du} + j(j+1) - \frac{l^2}{1 - u^2} \right] \chi_{j,l} = 0,$$

whose solutions are the associated Legendre functions, [**41**]. However, we can find their formulas directly. First, $\chi_{j,j}$ is an eigenfunction for J_+ with eigenvalue j. In spherical coordinates,

$$J_\pm = e^{\pm\,\mathrm{i}\,\theta} \left(\pm\frac{\partial}{\partial \varphi} + \mathrm{i}\,\cot \varphi\,\frac{\partial}{\partial \theta} \right),$$

and so $\chi_{j,j}$ is a solution to the first order ordinary differential equation

$$\frac{df}{d\varphi} - j\,(\cot \varphi)f = 0.$$

Solving, we find

$$\chi_{j,j}(\varphi) = c_j \sin^j \varphi, \qquad \psi_{j,j}(\varphi, \theta) = c_j e^{\,\mathrm{i}\, j\,\theta} \sin^j \varphi.$$

The other eigenfunctions are found by applying the lowering operator J_- repeatedly

$$\psi_{j,l}(\varphi, \theta) = J_-^{\,j-l}\,\psi_{j,j}(\varphi, \theta), \qquad l = -j,\ \ldots\ ,j.$$

This implies

$$\chi_{j,l}(\varphi) = c_{j,l} \sin^{-l} \varphi \, \frac{d^{j-l}}{d \cos \varphi^{j-l}} \sin^{2j} \varphi = c_{j,l} P_j^l(\cos \varphi), \qquad l = -j, \, \ldots \, , j,$$

where the P_j^l are the *associated Legendre functions*

$$P_j^l(u) = \frac{1}{2^j \, j!} \, (1 - u^2)^{l/2} \frac{d^{j+l}}{du^{j+l}} (u^2 - 1)^j. \tag{7.29}$$

For normalization, we take

$$\psi_{j,l}(\varphi, \theta) = Y_j^l(\varphi, \theta) = (-1)^l \sqrt{\frac{(2j+1)(j-l)!}{4\pi\,(j+l)!}} \, e^{\,\mathrm{i}\,l\,\theta} \, P_j^l(\cos \varphi), \tag{7.30}$$

where $Y_j^l(\varphi, \theta)$ is a certain trigonometric polynomial in $\cos \varphi$ and $\sin \varphi$, known as the $(j, l)^{\text{th}}$ *spherical harmonic*.

The reason that no half-integral representations appear in this particular example is a general property of the rotation group that is easily established.

Theorem 7.7. *There are no representations of* $\mathrm{SO}(3)$ *possessing half integral azimuthal quantum number* j.

Proof: Note that $\exp(2\pi\,\mathrm{i}\,J_z) = \mathrm{I}$, since $L_z = \mathrm{i}\,J_z$ generates the rotations around the z-axis. Therefore $\rho\big[\exp(2\pi\,\mathrm{i}\,J_z)\big]^j \psi = \psi$ for all ψ in the representation space. On the other hand,

$$\rho[\exp(2\pi\,\mathrm{i}\,J_z)]^j \psi = \exp(2\pi\,\mathrm{i}\,j)\,\psi,$$

hence j must be an integer. $\hspace{6cm}$ Q.E.D.

The problem is that the group $\mathrm{SO}(3)$ is not simply connected. Its simply connected covering group is the *special unitary group*

$$\mathrm{SU}(2) = \big\{\, A \in \mathrm{GL}(2, \mathbb{C}) \;\big|\; A^T \overline{A} = I, \;\; \det A = 1 \,\big\}. \tag{7.31}$$

Every matrix in $\mathrm{SU}(2)$ can be written in the form

$$A = \begin{pmatrix} \alpha & \beta \\ -\overline{\beta} & \overline{\alpha} \end{pmatrix}, \qquad \text{where} \qquad |\,\alpha\,|^2 + |\,\beta\,|^2 = 1.$$

Therefore, topologically, we can identify $\mathrm{SU}(2)$ with the three sphere $S^3 \subset \mathbb{C}^2$, which is compact and simply connected. If we write $\alpha = s + \mathrm{i}\,t, \;\; \beta = u + \mathrm{i}\,v$, then

$$A = s\,\mathrm{I} + t\,\mathrm{i}\,\sigma_z + u\,\mathrm{i}\,\sigma_y + v\,\mathrm{i}\,\sigma_x, \qquad s^2 + t^2 + u^2 + v^2 = 1,$$

where I is the identity matrix, and

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{7.32}$$

are the so-called *Pauli matrices*. The Lie algebra of SU(2), denoted $\mathfrak{su}(2)$, consists of all trace-free skew-adjoint (skew-Hermitian) 2×2 matrices

$$\mathfrak{su}(2) = \left\{ A \in \mathfrak{gl}(2, \mathbb{C}) \mid S^T = -\overline{S},\ \mathrm{tr}\, S = 0 \right\}. \tag{7.33}$$

Thus, its general element has the form

$$S = \begin{pmatrix} \mathrm{i}\, z & \mathrm{i}\, x - y \\ \mathrm{i}\, x + y & -\mathrm{i}\, z \end{pmatrix} = \mathrm{i}\, x\, \sigma_x + \mathrm{i}\, y\, \sigma_y + \mathrm{i}\, z\, \sigma_z,$$

which implies that the matrices $\mathrm{i}\,\sigma_x$, $\mathrm{i}\,\sigma_y$, $\mathrm{i}\,\sigma_z$ form a basis for $\mathfrak{su}(2)$. In calculations, it is slightly preferable to use the scaled basis

$$L_x = -\frac{\mathrm{i}}{2}\, \sigma_x, \qquad L_y = -\frac{\mathrm{i}}{2}\, \sigma_y, \qquad L_z = -\frac{\mathrm{i}}{2}\, \sigma_z, \tag{7.34}$$

which have the same commutation relations (7.13) as the Lie algebra $\mathfrak{so}(3)$. Therefore,

$$\mathfrak{so}(3) \simeq \mathfrak{su}(2)$$

are isomorphic Lie algebras, and hence SO(3) and SU(2) are locally isomorphic Lie groups.

To exhibit the isomorphism explicitly, consider the *adjoint representation* of SU(2) on its Lie algebra $\mathfrak{su}(2) \simeq \mathbb{R}^3$:

$$\rho(A)\, S = A\, S\, A^{-1}, \qquad S \in \mathfrak{su}(2), \qquad A \in \mathrm{SU}(2). \tag{7.35}$$

(The adjoint representation is a general construction of importance in the theory of Lie groups. The reader should check that it defines a valid representation.) Now

$$\det\left[\, \rho(A)\, S \,\right] = \det S = x^2 + y^2 + z^2,$$

and hence $\rho(A)$ is a real orthogonal transformation on $\mathfrak{su}(2) \simeq \mathbb{R}^3$. Moreover, as $\rho(A)$ is continuous and SU(2) is connected, $\det \rho(A) = 1$ for all $A \in \mathrm{SU}(2)$, so

$$\rho : \mathrm{SU}(2) \ \longrightarrow\ \mathrm{SO}(3).$$

It is easy to see that ρ covers SO(3) since ρ reduces to the identity on the Lie algebra level, and every matrix in SO(3) can be written as a product of exponentials. Moreover, $\rho(A) = \rho(-A)$, so ρ covers each rotation matrix twice. In fact, it can be readily checked that $\rho(A) = I$ if and only if $A = \pm\, I$, so ρ is a double covering and

$$\mathrm{SO}(3) \simeq \mathrm{SU}(2)/\mathbb{Z}_2, \tag{7.36}$$

where $\mathbb{Z}_2$ denotes the discrete subgroup consisting of $\pm\, I$. Therefore, as SU(2) was topologically a three-sphere, SO(3) is topologically the real projective space $\mathbb{RP}^3$.

The standard representation of the Lie algebra $\mathfrak{su}(2) \simeq \mathfrak{so}(3)$ is two-dimensional, and hence must be the half-integral representation corresponding to $j = \frac{1}{2}$. Indeed,

$$J^2 = -\tfrac{1}{4} s_1^2 - \tfrac{1}{4} s_2^2 - \tfrac{1}{4} s_3^2 = -\tfrac{3}{4}\, I,$$

which implies that $j(j+1) = \frac{3}{4}$, and hence $j = \frac{1}{2}$. In fact, the representations of $\mathfrak{so}(3)$ can all be written as symmetric powers of the basic representation of $\mathfrak{su}(2)$, i.e., if we view the

2×2 matrices in $\mathfrak{su}(2)$ acting on the vector space of linear polynomials in u, v, then the representation ρ_j will act on the homogeneous polynomials of degree $2j$ in u, v. Given

$$P(u, v) = a_0 u^{2j} + a_1 u^{2j-1} v + a_2 u^{2j-2} v^2 + \cdots + a_{2j} v^{2j},$$

then $Q = \rho_j(A)P$, where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SU}(2)$ is given by

$$Q(u, v) = P(a\,u + c\,v, b\,u + d\,v).$$

Indeed, the representation has dimension $j + 1$, and is irreducible, since the monomials $u^k v^{2j-k}$ form (up to scale) the basis constructed above:

$$\rho_j(J_+)P(u, v) = \frac{d}{dt} \rho_j(\exp t\, J_+)P(u, v) \bigg|_{t=0} = \frac{d}{dt} P(u, t\,u + v) \bigg|_{t=0} = u \frac{\partial P}{\partial v}.$$

Similarly, we deduce

$$\rho_j(J_+)P = u \frac{\partial P}{\partial v}, \qquad \rho_j(J_-)P = v \frac{\partial P}{\partial v}, \qquad \rho_j(J_z)P = \frac{u}{2} \frac{\partial P}{\partial u} + \frac{v}{2} \frac{\partial P}{\partial v}, \qquad (7.37)$$

and so, as required,

$$\rho_j(J_+)\,u^k v^{2j-k} = (2j - k)\,u^{k+1} v^{2j-k-1}, \qquad \rho_j(J_z)\,u^k v^{2j-k} = (k - j)\,u^k v^{2j-k},$$

$$\rho_j(J_-)\,u^k v^{2j-k} = k\,u^{k-1} v^{2j-k+1}, \qquad \rho_j(J^2)\,u^k v^{2j-k} = j(j + 1)\,u^k v^{2j-k}.$$

There is one remaining question: are these the only representations which appear? This is related to the issue of *completeness*, which is answered by the following result:

Theorem 7.8. *The spherical harmonics* $Y_j^l(\varphi, \theta)$, $j = 0, 1, 2, \ldots$, $l = -j, \ldots, j$, *form a basis for the space* $\mathrm{L}^2(S^2)$ *of square integrable functions on the sphere.*

In other words, every square-integrable function $f(\varphi, \theta)$ can be expanded in a "Fourier series" in terms of the spherical harmonics, which converges in norm. This theorem is a consequence of the celebrated Peter-Weyl Theorem, which states that the matrix elements of the irreducible representations of a compact Lie group G form a complete orthonormal basis for the space $\mathrm{L}^2(G)$ of square-integrable functions on G. The simplest example is the group $G = \mathrm{SO}(2)$, where these functions are $e^{i n \theta}$, and we are back to the theory of Fourier series. The spherical harmonics constitute the next important example of this general theory.

However, there is a fairly elementary direct proof of Theorem 7.8, cf. [**49**]. Consider the space of homogeneous harmonic polynomials of degree j:

$$\mathcal{P}_j = \left\{\, p(x, y, z) \,\middle|\, \Delta p = 0, \quad p(\lambda x, \lambda y, \lambda z) = \lambda^j p(x, y, z) \,\right\}.$$

If we write a polynomial p in spherical coordinates, we see that it has the form

$$p(r, \varphi, \theta) = r^j\, h(\varphi, \theta),$$

where, according to the formula for separating Laplace's equation in spherical coordinates

$$J^2 h = j(j+1)\, h,$$

where J^2 is the *spherical Laplacian* (7.28), and the $j(j+1)$ comes from the radial component of the Laplacian operator:

$$\Delta = \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} \right) - \frac{1}{r^2} J^2. \tag{7.38}$$

We conclude that h is a linear combination of spherical harmonics, and so a basis for $\mathcal{P}_j$ is provided by the basic harmonic polynomials

$$p(x, y, z) = r^j\, Y_j^l(\varphi, \theta), \qquad l = -j,\ \ldots\ , j.$$

To establish completeness, let $g(\varphi, \theta)$ be any continuous function on the unit sphere. We define

$$f(x, y, z) = r\, g(\varphi, \theta),$$

so f is continuous on all of $\mathbb{R}^3$. Now, by a Theorem of Weierstrass, we can approximate f uniformly by polynomials over any compact domain, which we assume contains the unit sphere. To show that g can be approximated by spherical harmonics, then, we need only show that any polynomial can be expressed as a sum of harmonic polynomials multiplied by powers of r. Specifically, if $P(x, y, z)$ is any polynomial of degree $n = 2m$ or $2m + 1$, then there exist harmonic polynomials $p_n, p_{n-2},\ \cdots\ , p_{n-2m}$, indexed by their degrees, such that

$$P(x, y, z) = p_n(x, y, z) + r^2\, p_{n-2}(x, y, z) + r^4\, p_{n-4}(x, y, z) + \cdots + r^{2m}\, p_{n-2m}(x, y, z). \tag{7.39}$$

We prove this claim by induction on the degree n. The result is obvious for $n = 0$ or 1 since every constant or linear polynomial is automatically harmonic. Suppose P has degree n. Then ΔP has degree $n - 2$, hence, by induction,

$$\Delta P(x, y, z) = q_{n-2}(x, y, z) + r^2\, q_{n-4}(x, y, z) + \ \cdots\ + r^{2m-2}\, q_{n-2m}(x, y, z),$$

where the q_k are harmonic polynomials. Set $p_k = q_k / [(n - k)(n + k + 1)]$, so that

$$\Delta[r^{2k}\, p_{n-2k}] = q_{n-2k}.$$

Then, if we define p_n by (7.39), we compute ΔP to find $\Delta p_n = 0$. *Q.E.D.*

The Hydrogen Atom

The simplest system with a Coulomb interaction is the hydrogen atom, which consists of a single proton and a single electron. By going to center of mass coordinates, the Hamiltonian reduces to

$$\mathcal{H} = -\frac{\hbar^2}{2m}\Delta + V(r) = -\frac{\hbar^2}{2m}\Delta - \frac{e^2}{r} \tag{7.40}$$

where we are considering the case of an attractive potential between two oppositely charged particles of charge $\pm e$. The stationary Schrödinger equation is then

$$\frac{\hbar^2}{2m}\Delta\psi + \left(\frac{e^2}{r} + E\right)\psi = 0, \tag{7.41}$$

where we interpret the eigenvalue as the energy: $\lambda = E$. Equation (7.41) can be solved by separation of variables in spherical coordinates r, φ, θ. We write

$$\psi(r, \varphi, \theta) = \xi(r)\,\chi(\varphi, \theta).$$

The Laplacian separates into

$$-\hbar^2\,\Delta = P^2 = P_r^2 + \frac{1}{r^2}\,L^2,$$

where

$$P_r = -\mathrm{i}\,\hbar\,\frac{1}{r}\,\frac{\partial}{\partial r}\,r = -\mathrm{i}\,\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right)$$

is the radial momentum operator, so

$$P_r^2 = -\hbar^2\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right),$$

whereas

$$L^2 = \hbar^2 J^2 = -\hbar^2\left(\frac{1}{\sin\varphi}\frac{\partial}{\partial\varphi}\sin\varphi\frac{\partial}{\partial\varphi} + \frac{1}{\sin^2\varphi}\frac{\partial^2}{\partial^2\theta}\right)$$

is the total moment of angular momentum:

$$L^2 = L_x^2 + L_y^2 + L_z^2, \qquad \mathbf{L} = (L_x, L_y, L_z) = -\mathrm{i}\,\hbar\,\mathbf{q}\wedge\nabla. \tag{7.42}$$

For a general spherically symmetric potential, then, the stationary Schrödinger equation separates into the angular component

$$L^2\chi = \mu\chi, \tag{7.43}$$

where μ is the separation constant, and a radial component

$$\frac{1}{2m}\left(P_r^2 + \frac{\mu}{r^2}\right)\xi + V(r)\xi = E\,\xi,$$

which is the same as

$$\left[\frac{\hbar^2}{2m}\left(-\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{\mu}{r^2}\right) + V(r) - E\right]\xi = 0.$$

If we set

$$\eta(r) = r\,\xi(r),$$

then the equation becomes

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2\mu}{2mr^2} + V(r) - E\right]\eta = 0. \tag{7.44}$$

Note the resemblance with the usual one-dimensional Schrödinger equation (6.15).

The angular equation (7.43) does not depend on the particular potential under consideration. Since L^2 is just a multiple of the spherical Laplacian (7.28), its solutions are the standard spherical harmonics $Y_j^l(\varphi, \theta)$, as in (7.30). Therefore, the separation constant μ only achieves the values $j(j+1)$, where j, the azimuthal quantum number, is integral, corresponding to the integral representations of SO(3).

In the case of a Coulomb interaction, $V(r) = -e^2/r$, and the radial equation (7.44) becomes

$$\eta'' + \left[-\frac{j(j+1)}{r^2} + \frac{2\,m\,e^2}{\hbar^2 r} + \frac{2\,m\,E}{\hbar^2} \right] \eta = 0. \tag{7.45}$$

If the energy $E > 0$, then the solution to this equation is oscillatory, and represents an unbound scattering state, assocaited with the continuous spectrum of the Schrödinger operator. If $E < 0$, then the asymptotic form of a solution which is regular at the origin will be a linear combination of exponentials $e^{\kappa r}$ and $e^{-\kappa r}$, where $\kappa = \sqrt{-2\,m\,E}/\hbar$. For this to be an acceptable eigenfunction, the coefficient of the increasing exponential must vanish; this will only happen for certain discrete values of the energy E.

If we make the change of variables

$$x = 2\kappa r = \frac{2r\sqrt{-2\,m\,E}}{\hbar},$$

then the radial equation reduces to one only involving the dimensionless parameter

$$\nu = \frac{e^2}{\hbar}\sqrt{\frac{m}{-2\,E}},$$

namely

$$\eta'' + \left[-\frac{j(j+1)}{x^2} + \frac{\nu}{x} - \frac{1}{4} \right] \eta = 0. \tag{7.46}$$

If we make the change

$$\eta(x) = x^{j+1} e^{-x/2} \zeta(x),$$

then ζ satisfies the *Laplace ordinary differential equation* (not to be confused with the Laplace partial differential equation!)

$$x\,\zeta'' + (2\,j + 2 - x)\,\zeta' - (j + 1 - \nu)\,\zeta = 0. \tag{7.47}$$

The solution which is regular at the origin can be written in terms of the *confluent hypergeometric function*

$$\zeta(x) = F(j+1-\nu, 2\,j+2, x) = \sum_{i=0}^{\infty} \frac{\Gamma(j+1+i-\nu)\,(2\,j+1)!\,x^i}{\Gamma(j+1-\nu)\,(2\,j+1+i)!\,i!}. \tag{7.48}$$

In general, this represents an infinite series, asymptotically behaving as $x^{-j-1-\nu} e^x$ for x large, and hence cannot represent an eigenfunction. However, if $j + 1 - \nu$ is a negative integer, then the series terminates, and the solution reduces to a polynomial, and we have

an eigenfunction. Thus the bound states of the hydrogen atom are prescribed by the quantum condition

$$\nu = n = j + k + 1, \qquad \text{where} \qquad k = 0, 1, 2, \ldots,$$

is the *radial quantum number*. The solutions have the form

$$x^{j+1} e^{-x/2} L_k^{2j+1}(x), \tag{7.49}$$

where L_k^i denotes the *associated Laguerre polynomial*:

$$L_k^0(x) = e^x \frac{d^k}{dx^k} x^k e^{-x}, \qquad L_k^i(x) = (-1)^i \frac{d^i}{dx^i} L_{i+k}^0(x). \tag{7.50}$$

They represent orthogonal polynomials on the interval $(0, \infty)$ with respect to the weight function $x_k e^{-x}$. Moreover, L_k^i is a polynomial of degree k having exactly k zeros on the interval $(0, \infty)$.

If we replace the parameter ν by its assumed value in terms of the angular momentum j and the radial quantum number k, and solve for the energy E, we deduce the complete energy spectrum of the hydrogen atom:

$$E_{j,k} = -\frac{m\,e^4}{2\,\hbar\,(j+k+1)^2}. \tag{7.51}$$

Therefore, the energy takes on a denumerably infinite number of values, since both the azimuthal and radial quantum numbers can take on any non-negative integral value. The energy itself only depends on the *principal quantum number*

$$n = j + k + 1, \tag{7.52}$$

with the energy level corresponding to

$$E_n = -\frac{m\,e^4}{2\,\hbar^2\,n^2} = -\frac{\alpha^2\,\varepsilon}{2\,n^2}, \tag{7.53}$$

where

$$\varepsilon = m\,c^2 = .51072 \text{ MeV} \qquad (1 \text{ MeV} = 10^6 \text{ electron volts}) \tag{7.54}$$

is the rest energy of the electron (m is actually the reduced mass), and the dimensionless number

$$\alpha = \frac{e^2}{\hbar\,c} \approx \frac{1}{137}, \tag{7.55}$$

known as the *fine structure constant*. Numerically, the ground state energy, $n = 1$, has the value $E_1 = -13.5983 \text{ eV}$.

The energy level E_n is degenerate, since the angular momentum quantum number j can assume values from 0 up to $n - 1$; the order of degeneracy is

$$\sum_{j=0}^{n-1} (2j + 1) = n^2.$$

In traditional spectroscopy, the various eigenstates are designated by the positive integer n followed by a letter, $s, p, d, f, g, \ldots$, indicating the value of $j = 0, 1, 2, \ldots$. The magnetic quantum number l appearing in the spherical harmonic component, indicating the orientation of the state, is simply not mentioned. Thus we have the following states:

Energy	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$\ldots$
E_1	$1\,s$				
E_2	$2\,s$	$2\,p$			
E_3	$3\,s$	$3\,p$	$3\,d$		
E_4	$4\,s$	$4\,p$	$4\,d$	$4\,f$	

The energy levels (7.53) accumulate at 0, after which the continuous spectrum starts. (This qualitative behavior is common to all "short-range potentials".) The normalized eigenfunctions corresponding to the quantum state labelled by n, j, l is

$$\psi_{n,j,l}(x) = \frac{2^{j+1}}{n^{j+2}\, a^{j+(3/2)}} \sqrt{\frac{(n-j-1)!}{(n+j)!^3)}} \; r^j L_{n-j-1}^{2j+1}\left(\frac{2r}{na}\right) e^{-r/(na)} Y_j^l(\varphi, \theta), \qquad (7.56)$$

where

$$a = \frac{\hbar^2}{m\,e^2} = 0.529 \times 10^{-8} \text{ cm}, \qquad (7.57)$$

This number, which has the units of distance, is known as the radius of the hydrogen atom, or the *Bohr radius*.

If the system is in the pure state $\psi_{n,j,l}$, the mean value of the radius r gives the average radial distance of the electron from the nucleus. A calculation proves that the expected value of the radius in the eigenstate (7.56) is

$$\langle r \rangle_{n,j,l} = \frac{3n^2 - j(j+1)}{2}\, a, \qquad (7.58)$$

independent of l. In particular, $\frac{3}{2}\,a$ is the mean value of the radius of a hydrogen atom in the ground state $(n = 1, j = 0)$. Note that the radii are not monotone in n, so states of higher angular momentum and higher energy may lie "closer" to the nucleus than ones of lower energy. (However, this only happens for $n \geq 7$, which is rare in physical situations.) We also have

$$\langle r^2 \rangle_{n,j,l} = \frac{n^2 \left[5n^2 + 1 - 3j(j+1)\right]}{2}\, a^2,$$

When $j = n - 1$ is maximal, the wave function is especially simple. We find

$$\langle r \rangle_n = n^2\left(n + \tfrac{1}{2}\right)a, \qquad \langle r^2 \rangle_n = n^2\left(n + \tfrac{1}{2}\right)(n+1)a^2,$$

hence the root-mean square deviation in r is

$$\Delta r_n = \sqrt{\langle r^2 \rangle_n - \langle r \rangle_n^2} = \tfrac{1}{2}\,n\sqrt{2n+1}\,a = \frac{\langle r \rangle_n}{\sqrt{2n+1}}.$$

Thus, for large values of n, $\Delta r_n / \langle r \rangle_n$ becomes small, and the electron remains practically localized in a sphere of radius $n^2 a$, whereas the energy of the level, $-e^2/n^2 a$, is the same as that of a classical electron describing a circular orbit or radius $n^2 a$. This gives an illustration of the general correspondence rule that the quantum system reduces to the classical laws of motion for very large quantum numbers. The states of maximal angular momentum correspond to the classical circular orbits.

The Zeeman Effect

In the physical measurements of a quantum system, we can never detect the degeneracies of a given energy level directly without some physical operator to distinguish the wave functions for that level. The magnetic quantum number of the basic hydrogen atom needs some additional physical effect to make itself manifest. This is provided by the Zeeman effect, obtained by measuring the spectrum of the hydrogen atom in a magnetic field. The cylindrically symmetric magnetic field has the effect of breaking the spherical symmetry of the atom and reducing the symmetry group from SO(3) to SO(2). Each representation of SO(3) corresponding to the azimuthal quantum number j then splits up into $2j + 1$ distinct one-dimensional representations of SO(2).

We begin with the classical Hamiltonian for a hydrogen atom:

$$H = \frac{|\mathbf{p}|^2}{2m} - \frac{e^2}{r}.$$

The Hamiltonian of the same atom placed in a static magnetic field with potential $\mathbf{A}(\mathbf{q})$ is given by replacing the momentum $\mathbf{p}$ by $\mathbf{p} - e\mathbf{A}(\mathbf{q})/c$. In particular, for a constant magnetic field, $\mathbf{A} = \frac{1}{2}\mathbf{H} \wedge \mathbf{q}$, and we have

$$\left| \mathbf{p} - \frac{e}{c}\mathbf{A}(\mathbf{q}) \right|^2 = |\mathbf{p}|^2 - \frac{e}{c}\mathbf{H} \cdot \mathbf{L} + \frac{e^2}{4c^2}|\mathbf{H}|^2 \cdot |\mathbf{q}_\perp|^2,$$

where $\mathbf{q}_\perp$ denotes the projection of $\mathbf{q}$ onto the plane perpendicular to the field $\mathbf{H}$. Therefore, the corresponding quantum Hamiltonian takes the form

$$\mathcal{H} = \mathcal{H}_0 - \frac{e}{2mc}\mathbf{H} \cdot \mathbf{L} + \frac{e^2}{8mc^2}|\mathbf{H}|^2|\mathbf{q}_\perp|^2,$$

where $\mathcal{H}_0$ denotes the Hamiltonian for the free atom, and $\mathbf{L}$ denotes the usual angular momentum operator (7.42). In the present situation, the third term plays a negligible role, (even for a very strong field and a very heavy atom) and can be ignored. Thus, we consider the reduced Hamiltonian

$$\mathcal{H} = \mathcal{H}_0 - \frac{e}{2mc}\mathbf{H} \cdot \mathbf{L}. \tag{7.59}$$

Now, the eigenfunctions $\psi_{n,j,l}$ of $\mathcal{H}_0$ have energy eigenvalue $E_{n,j}$. Actually, for a hydrogen atom, accidental symmetry implies that the energy $E_{n,j} = E_n$ only depends on the principal quantum number n, cf. (7.53). This is no longer the case when dealing with multiple electron atoms, where one can distinguish the energy levels associated with differing azimuthal quantum numbers j. Since the ensuing argument remains valid, we retain the

dependence of the energy on both n and j. If we assume $\mathbf{H} = h_0 \mathbf{k}$ is in the direction of the z-axis, then these remain eigenfunctions of the modified Hamiltonian, but with eigenvalues

$$E_{n,j} - \mu h_0 l = -\frac{m e^4}{2 \hbar n^2} - \mu h_0 l, \qquad l = -j, \; \ldots \; , j, \tag{7.60}$$

where the constant

$$\mu = \frac{e \hbar}{2 m c} = 0.927 \times 10^{-20} \text{ erg/gauss} \tag{7.61}$$

is known as the *Bohr magneton*, and can be identified as the magnetic moment of an electron. Therefore, each degenerate level has split into an odd "multiplet" of $2j + 1$ equally spaced levels, whose spacing is independent of the atom under consideration, and whose average energy is $E_{n,j}$. These theoretical predictions are only partly confirmed by experiment. For atoms with an odd number of electrons, the multiplets are all even (the anomalous Zeeman effect), which is what one would expect if the azimuthal quantum number were a half integer. Also the spacing of levels varies from multiplet to multiplet. These are all handled by the hypothesis of electron spin, to be discussed below.

Addition of Angular Momenta

Suppose we have a system consisting of several particles. The total angular momentum of the system will be a sum of the individual angular momenta,

$$J = \sum J_\nu.$$

The underlying representation space will be the tensor product of the individual Hilbert spaces. Indeed, by separation of variables, we can write the general wave function as a sum of products of wave functions for the individual subsystems.

Let G be a group. By the *tensor product* of two representations ρ_1, ρ_2, we mean the representation

$$\rho = \rho_1 \otimes \rho_2 \colon G \; \longrightarrow \; \mathcal{L}(V_1 \otimes V_2), \qquad \rho(g) = \rho_1(g) \otimes \rho_2(g), \tag{7.62}$$

on the tensor product vector space. The matrix entries of the tensor product are the products of the corresponding entries of the individual representations. If the representations are unitary, then we can decompose the tensor product into irreducible representations

$$\rho_1 \otimes \rho_2 = \bigoplus_k \rho_k,$$

where the factors may appear with various multiplicities. In particular, the tensor product of irreducible representations is not necessarily irreducible, and we have formulae of the form

$$\rho_i \otimes \rho_j = \bigoplus_k a_{i,j}^k \, \rho_k, \tag{7.63}$$

known as the *Clebsch-Gordan series* (although in finite dimensions, it is necessarily a finite sum).

Remark: This is a bit of a misnomer, since Clebsch and Gordan only worked in classical invariant theory, and didn't have representation theory in mind at all. It appears that the terminology comes from a remark in Weyl 's book [**54**] in which he says that this series is "... essentially identical with the Clebsch-Gordan series in ... [classical] invariant theory."

In order to compute the Clebsch-Gordan series, we need to know how to recognize which irreducible representations occur in a general reducible representation. The simplest approach to this is through the theory of group characters, which is of importance in its own right.

Definition 7.9. Let $\rho\colon G \longrightarrow \mathcal{L}(V)$ be a representation of the group G on a finite-dimensional vector space V. The *character* of the representation ρ is the map obtained by taking the trace:

$$\chi\colon G \longrightarrow \mathbb{C}, \qquad \chi(g) = \operatorname{tr} \rho(g). \tag{7.64}$$

Note that since $\operatorname{tr} S^{-1}AS = \operatorname{tr} A$, the character of a representation is independent of any basis on V, and so equivalent representations have the same character. For finite or compact Lie groups, the converse also holds, as we will see below. Also, if the representation is unitary, then

$$\chi(g^{-1}) = \operatorname{tr} \rho(g^{-1}) = \operatorname{tr} \rho(g)^{-1} = \operatorname{tr} \rho(g)^* = \overline{\chi(g)}.$$

Given a representation ρ of a compact Lie group or finite group G and a basis $\{e_1, \ldots, e_n\}$ of the representation space V, we define the corresponding matrix elements

$$\rho_{ij}(g), \qquad i,j = 1, \ldots, n,$$

which are continuous functions on the group G. We define an inner product on G by using the invariant integral

$$\langle\, f\,, h\,\rangle = \int_G f(g)\, h(g)^*\, dg \tag{7.65}$$

given by the Haar measure on G to average over the group. (This is normalized so that the total measure of G is 1.) For a finite group, we use the analogous averaging sum.

Theorem 7.10. *Let ρ^λ and ρ^μ be irreducible representations of the compact Lie or finite group G. Then their matrix elements satisfy the following orthogonality relations:*

$$\langle\, \rho_{ij}^\lambda\,, \rho_{kl}^\mu\,\rangle = 0, \quad \text{unless} \quad \lambda = \mu, \quad i = j, \quad k = l,$$
$$\langle\, \rho_{ij}^\lambda\,, \rho_{ij}^\lambda\,\rangle = \frac{1}{\dim V_\lambda}. \tag{7.66}$$

Proof: Let $B\colon V^\mu \longrightarrow V^\lambda$ be any linear transformation. Define $A\colon V^\mu \longrightarrow V^\lambda$ by

$$A = \int_G \rho^\lambda(g)\, B\, \rho^\mu(g)^{-1} dg = \int_G \rho^\lambda(g)\, B\, \rho^\mu(g)^* dg. \tag{7.67}$$

The invariance of the integral readily implies that

$$\rho^\mu(g)\, A = A\, \rho^\lambda(g)$$

for any $g \in G$. Therefore, according to Schur's Lemma 7.5, $A = 0$ if $\lambda \neq \mu$, while $A = \alpha\,\mathrm{I}$ if $\lambda = \mu$. Choosing B to have all zero entries except for a single 1, the last integral in (7.67) reduces to

$$A_{ik} = \int_G \rho_{ij}^\lambda(g)\,\rho_{kl}^\mu(g)^*\,dg = \langle\,\rho_{ij}^\lambda\,,\rho_{kl}^\mu\,\rangle.$$

This implies the first set of zero orthogonality relations. To check the latter set, we need to compute α, which is done by taking traces:

$$\alpha \dim V^\lambda = \mathrm{tr}(\alpha\,\mathrm{I}) = \mathrm{tr}\,A = \int_G \mathrm{tr}\,\rho^\lambda(g)\,B\,\rho^\lambda(g)^{-1}dg = \int_G \mathrm{tr}\,B\,dg = \mathrm{tr}\,B.$$

Taking B as before completes the proof. $\hspace{8cm}$ Q.E.D.

The matrix elements of the irreducible representations of standard Lie groups can be expressed in terms of the special functions of mathematical physics, e.g., spherical harmonics, Bessel functions, hypergeometric functions, etc. Many of the fancy addition formulas, recurrence relations, differential equations, etc. satisfied by these functions are then simple by-products of standard representation theory. See the book by Talman, [**49**], for an introduction to this "unified theory of special functions".

Corollary 7.11. *The characters χ^λ of the irreducible representations ρ^λ of a finite or compact Lie group G satisfy the orthogonality relations:*

$$\langle\,\chi^\lambda\,,\chi^\mu\,\rangle = \delta_{\lambda,\mu}. \tag{7.68}$$

Theorem 7.12. *. Let ρ be an arbitrary finite-dimensional unitary representation of G with character χ. Then ρ decomposes as a direct sum of irreducible representations*

$$\rho = \bigoplus_\lambda n_\lambda \rho^\lambda.$$

The number of times the irreducible representation ρ^λ occurs in ρ is given by

$$n_\lambda = \langle\,\chi\,,\chi^\lambda\,\rangle.$$

Proof: Since χ is a trace, and the ρ^λ occur in block diagonal form in ρ, we have

$$\chi = \sum_\lambda n_\lambda \chi^\lambda.$$

The result then follows from the orthogonality relations (7.68). $\hspace{5cm}$ Q.E.D.

Theorem 7.12 can be used to compute the Clebsch-Gordan coefficients for tensor products of representations; we need only determine the character of the tensor product, and rewrite it in terms of the characters of the irreducible representations. Note that the character of the tensor product of two representations is just the ordinary product of the two characters:

$$\chi^1 \otimes \chi^2(g) = \chi^1(g)\,\chi^2(g).$$

Therefore, to determine the Clebsch-Gordan series of the tensor product of two irreducible representations, we need only decompose the product of their characters into a sum of irreducible characters:

$$\chi^i \chi^j = \bigoplus_k a_{i,j}^k \chi^k,$$

where, according to the previous result

$$a_{i,j}^k = \langle \chi^i \chi^j, \chi^k \rangle.$$

In practice, though, it is often easier to determine the Clebsch-Gordan series directly from the character expansion.

In the case of SU(2), we proceed as follows. (The same arguments apply to the orthogonal group SO(3).) Consider the irreducible representation ρ^j on the space of homogeneous polynomials of degree $2j$, as described above. The character χ^j of ρ^j can be computed explicitly by taking the trace relative to the standard basis $u^k v^{2j-k}$, but this expression is too complicated to make practical use of. A better way to proceed is as follows. Let $A \in$ SU(2) be diagonalized by a similarity matrix

$$B = SAS^{-1} = \begin{pmatrix} e^{i\tau/2} & 0 \\ 0 & e^{-i\tau/2} \end{pmatrix}, \qquad -2\pi < \tau \leq 2\pi.$$

(Note that τ represents the angle of the corresponding rotation in SO(3).) Then the representation of B is diagonal:

$$\rho^j(B) u^k v^{2j-k} = e^{i(j-k)\tau} u^k v^{2j-k}.$$

Therefore

$$\chi^j(A) = \chi^j(B) = \operatorname{tr} \rho^j(B) = \sum_{k=0}^{2j} e^{i(j-k)\tau} = \sum_{k=-j}^{j} e^{ik\tau} = \frac{\sin\left(j + \frac{1}{2}\right)\tau}{\sin \frac{1}{2}\tau}.$$

Therefore, to determine the Clebsch-Gordan series, we need to rewrite the product $\chi^i \chi^j$ in terms of the χ^k. We compute, assuming $j \geq i$:

$$\chi^i \otimes \chi^j(A) = \chi^i(A)\,\chi^j(A) = \sum_{p=-i}^{i} \sum_{q=-j}^{j} e^{ip\tau} e^{iq\tau} = \sum_{k=j-i}^{j+i} \sum_{n=-k}^{k} e^{ik\tau} = \sum_{k=j-i}^{j+i} \chi^k(A).$$

Therefore, we conclude:

$$\rho^i \otimes \rho^j = \rho^{i+j} \oplus \rho^{i+j-1} \oplus \cdots \oplus \rho^{|i-j|}. \tag{7.69}$$

Theorem 7.13. *The tensor product $\rho^i \otimes \rho^j$ of two irreducible representations of* SU(2) *decomposes as a direct sum of one copy of each of the irreducible representations ρ^k for $k = i+j, i+j-1, \ldots, |i-j|$.*

Corollary 7.14. *In a system of two particles, if the first particle has total angular momentum i and the second has total angular momentum j, then the system has possible total angular momenta ranging from $|i-j|$ to $i+j$.*

Note that this holds for both integral and half-integral values of i and j. In particular, two half-integral representations combine to give a direct sum of integral representations.

In fact, we can write this decomposition even more explicitly. Suppose we have introduced the basis vectors $\psi_{j,l}$ of the representation ρ_j as above. Then the tensor product of basis elements must decompose into a sum of basis elements of the component representations, i.e., we have

$$\psi_{i,l} \otimes \psi_{j,m} = \sum_{k=j-i}^{j+i} \sum_{n=-k}^{k} C_{i,j,l,m}^{k,n} \psi_{k,n}. \tag{7.70}$$

The coefficients $C_{i,j,l,m}^{k,n}$ are known as the *Wigner coefficients*, also known as *Clebsch-Gordan coefficients* or *vector coupling coefficients*. They are uniquely determined up to a phase factor, since the normalized eigenvectors $\psi_{j,m}$ are determined up to a phase. In fact, it possible to consistently choose the phases such that all the Wigner coefficients are real. Note first that by orthonormality, which is guaranteed by the unitarity of the representations,

$$C_{i,j,l,m}^{k,n} = \langle \psi_{i,l} \otimes \psi_{j,m}, \psi_{k,n} \rangle. \tag{7.71}$$

Hence, assuming reality — otherwise the coefficient would have a complex conjugate — we have the converse expansion

$$\psi_{k,n} = \sum_{i,j,l,m} C_{i,j,l,m}^{k,n} \psi_{i,l} \otimes \psi_{j,m}. \tag{7.72}$$

Next, note that

$$J_z \psi_{j,m} = m \psi_{j,m}, \qquad \text{hence} \qquad J_z(\psi_{i,l} \otimes \psi_{j,m}) = (l+m) \psi_{i,l} \otimes \psi_{j,m}.$$

Applying J_z to both sides of (7.72), we conclude that

$$C_{i,j,l,m}^{k,n} = 0 \qquad \text{unless} \qquad l + m = n.$$

A slick way to compute the Wigner coefficients is to use an invariant theoretic argument.

Theorem 7.15. *Let $(u,v), (x,y)$ and (ξ, η) denote independent variables. Define the function*

$$F(u,v,x,y,x,h) = (u\,y - v\,x)^{i+j-k}(\xi\,u + \eta\,v)^{i-j-k}(\xi\,x + \eta\,y)^{j-i-k}. \tag{7.73}$$

Then the coefficient of

$$\frac{\xi^{k+n} \eta^{k-n}}{\sqrt{(k+n)!\,(k-n)!}} \tag{7.74}$$

is given by a multiple, depending only on i, j, k, of

$$\sum_{i,j,l,m} C_{i,j,l,m}^{k,n} \frac{u^{i+l}\,v^{i-l}}{\sqrt{(i+l)!\,(i-l)!}} \frac{x^{j+m}\,y^{j-m}}{\sqrt{(j+m)!\,(j-m)!}}. \tag{7.75}$$

For the proof, one demonstrates that the function (7.73) is invariant under the group SU(2) acting covariantly on $\mathbf{u} = (u, v)$ and $\mathbf{x} = (x, y)$ and contravariantly on $\boldsymbol{\omega} = (\xi, \eta)$:

$$\mathbf{u} \longmapsto A\,\mathbf{u}, \qquad \mathbf{x} \longmapsto A\,\mathbf{x}, \qquad \boldsymbol{\omega} \longmapsto A^{-T}\boldsymbol{\omega}, \qquad A \in \mathrm{SU}(2).$$

This implies that the quantities

$$\mathbf{u} \wedge \mathbf{x} = uy - vx, \qquad \mathbf{u} \cdot \boldsymbol{\omega} = \xi u + \eta v, \qquad \mathbf{x} \cdot \boldsymbol{\omega} = \xi x + \eta y,$$

are invariant under SU(2), and, indeed, under all of SL(2, $\mathbb{C}$). On the other hand, the simpler function

$$(\xi\, u + \eta\, v)^{2j} = (2\,j)! \sum_n \frac{u^{k+n}\, v^{k-n}}{\sqrt{(k+n)!\,(k-n)!}}\, \frac{\xi^{k+n}\, \eta^{k-n}}{\sqrt{(k+n)!\,(k-n)!}}$$

is also clearly invariant under the same group. Therefore, the coefficients of (7.74) in each expression have the same covariant transformation rules, which suffices to prove the theorem. $\hfill$ Q.E.D.

Evaluating (7.73), we find a closed form expression for the Wigner coefficients.

Corollary 7.16. *For $l + m = n$, $C_{i,j,l,m}^{k,n}$ is equal to a multiple $\alpha_{i,j}^k$ of the sum*

$$\sum_p (-1)^p \frac{\sqrt{(i+l)!\,(i-l)!\,(j+m)!\,(j-m)!\,(k+n)!\,(k-n)!}}{(i-l-p)!\,(n+l-j+p)!\,(j+m-p)!\,(n-i-m+p)!\,p!\,(i+j-n-p)!}$$

(7.76)

The (finite) sum is over all integers p such that no negative factorials occur. To obtain normalized eigenvectors, one chooses

$$\alpha_{i,j}^k = \sqrt{2\,k+1}\,\sqrt{\frac{(k+i-j)!\,(k-i+j)!\,(i+j-k)!}{(i+j+k+1)!}}\,.$$

Another common quantity associated with the Wigner coefficients are the 3-*j symbols* defined by

$$C_{i,j,l,m}^{k,n} = (-1)^{i-j+n}\,\sqrt{2\,k+1}\,\begin{pmatrix} i & j & k \\ l & m & -n \end{pmatrix}.$$

The 3-*j* symbols are so called because i, j, k are often denoted j_1, j_2, j_3. Note that they are zero unless the sum of the three lower indices is 0. They are invariant under circular permutations of the columns, and get multiplied by $(-1)^{i+j+k}$ under transposition of two columns, or under simultaneous change of the signs of l, m, n.

Another convenient means of computing these coefficients is the "step down method". Note first that if we have maximal values for k, n, so that $i + j = k = n$, then there is only one term in the sum (7.76). Therefore, we can set

$$C_{i,j,i,j}^{i+j,i+j} = 1, \qquad \text{i.e.,} \qquad \psi_{i+j,i+j} = \psi_{i,i} \otimes \psi_{j,j}.$$

If we then apply J_- to this, we find, using formula (7.26),

$$J_-\psi_{i+j,i+j} = \sqrt{2i+2j}\,\psi_{i+j,i+j-1} = \sqrt{2i}\,\psi_{i,i-1}\otimes\psi_{j,j} + \sqrt{2j}\,\psi_{i,i}\otimes\psi_{j,j-1}.$$

Therefore

$$\psi_{i+j,i+j-1} = \sqrt{\frac{i}{i+j}}\,\psi_{i,i-1}\otimes\psi_{j,j} + \sqrt{\frac{j}{i+j}}\,\psi_{i,i}\otimes\psi_{j,j-1},$$

hence

$$C^{i+j,i+j-1}_{i,j,i-1,j} = \sqrt{\frac{i}{i+j}}, \qquad C^{i+j,i+j-1}_{i,j,i,j-1} = \sqrt{\frac{j}{i+j}}.$$

There is another eigenvector for J_z corresponding to the eigenvalue $i+j-1$, namely $\psi_{i+j-1,i+j-1}$. This will be orthogonal to $\psi_{i+j,i+j-1}$, and hence can be taken to be

$$\psi_{i+j-1,i+j-1} = \sqrt{\frac{j}{i+j}}\,\psi_{i,i-1}\otimes\psi_{j,j} - \sqrt{\frac{i}{i+j}}\,\psi_{i,i}\otimes\psi_{j,j-1}.$$

hence

$$C^{i+j-1,i+j-1}_{i,j,i-1,j} = \sqrt{\frac{j}{i+j}}, \qquad C^{i+j-1,i+j-1}_{i,j,i,j-1} = -\sqrt{\frac{i}{i+j}}.$$

One can continue this process of applying J_- and using orthogonality to deduce recurrence relations for the Wigner coefficients:

$$\sqrt{(k+n)(k-n+1)}\,C^{k,n+1}_{i,j,l,m}$$
$$= \sqrt{(i-l)(i+l+1)}\,C^{k,n}_{i,j-1,l,m} + \sqrt{(j-m)(j+m+1)}\,C^{k,n}_{i,j,l,m-1}.$$

A similar recurrence comes from applying J_+:

$$\sqrt{(k-n)(k+n+1)}\,C^{k,n-1}_{i,j,l,m}$$
$$= \sqrt{(i+l)(i-l+1)}\,C^{k,n}_{i,j+1,l,m} + \sqrt{(j+m)(j-m+1)}\,C^{k,n}_{i,j,l,m+1}.$$

When $k=n$, the left hand side of the first recurrence vanishes, which allows one to derive all the Wigner coefficients $C^{k,k}_{i,j,l,m}$ in terms of one of them, say $C^{k,k}_{i,j,i,j}$ say, which is given by the general formula above.

Example 7.17. For two particles of angular momentum $i = j = \frac{1}{2}$, each particle has two possible states corresponding to $m = \pm\frac{1}{2}$, which we denote by ψ_+, ψ_-, respectively. The combined system has angular momentum $k = 1$ or 0, with eigenfunctions ψ_0 for $k = 0$, and $\varphi_-, \varphi_0, \varphi_+$, corresponding to $k = 1$, $m = -1, 0, 1$, respectively. The Wigner coefficients for this simple case correspond to the rules

$$\varphi_+ = \psi_+\otimes\psi_+, \qquad \varphi_0 = \sqrt{\tfrac{1}{2}}\,\psi_+\otimes\psi_- + \sqrt{\tfrac{1}{2}}\,\psi_-\otimes\psi_+,$$

$$\varphi_- = \psi_-\otimes\psi_-, \qquad \psi_0 = \sqrt{\tfrac{1}{2}}\,\psi_+\otimes\psi_- - \sqrt{\tfrac{1}{2}}\,\psi_-\otimes\psi_+.$$

More complicated is the case of two particles of angular momentum 1. The combined system has total angular momentum $0, 1$, or 2, and we find:

$$\psi_{2,2} = \psi_{1,1} \otimes \psi_{1,1},$$

$$\psi_{2,1} = \sqrt{\tfrac{1}{2}}\, \psi_{1,1} \otimes \psi_{1,0} + \sqrt{\tfrac{1}{2}}\, \psi_{1,0} \otimes \psi_{1,1},$$

$$\psi_{1,1} = \sqrt{\tfrac{1}{2}}\, \psi_{1,1} \otimes \psi_{1,0} - \sqrt{\tfrac{1}{2}}\, \psi_{1,0} \otimes \psi_{1,1},$$

$$\psi_{2,0} = \sqrt{\tfrac{1}{6}}\, \psi_{1,1} \otimes \psi_{1,-1} + \sqrt{\tfrac{2}{3}}\, \psi_{1,0} \otimes \psi_{1,0} + \sqrt{\tfrac{1}{6}}\, \psi_{1,-1} \otimes \psi_{1,1},$$

$$\psi_{1,0} = \sqrt{\tfrac{1}{2}}\, \psi_{1,1} \otimes \psi_{1,-1} - \sqrt{\tfrac{1}{2}}\, \psi_{1,-1} \otimes \psi_{1,1},$$

$$\psi_{0,0} = \sqrt{\tfrac{1}{3}}\, \psi_{1,1} \otimes \psi_{1,-1} - \sqrt{\tfrac{1}{3}}\, \psi_{1,0} \otimes \psi_{1,0} + \sqrt{\tfrac{1}{3}}\, \psi_{1,-1} \otimes \psi_{1,1},$$

$$\psi_{2,-1} = \sqrt{\tfrac{1}{2}}\, \psi_{1,-1} \otimes \psi_{1,0} + \sqrt{\tfrac{1}{2}}\, \psi_{1,0} \otimes \psi_{1,-1},$$

$$\psi_{1,-1} = \sqrt{\tfrac{1}{2}}\, \psi_{1,-1} \otimes \psi_{1,0} - \sqrt{\tfrac{1}{2}}\, \psi_{1,0} \otimes \psi_{1,-1},$$

$$\psi_{2,-2} = \psi_{1,-1} \otimes \psi_{1,-1}.$$

The 3-j coefficients also arise naturally in the addition of three angular momenta. Specifically, suppose we have a system of three particles, with azimuthal quantum numbers j_1, j_2, j_3 and magnetic quantum numbers l_1, l_2, l_3. For the system to have zero total angular momentum, we must have $l_1 + l_2 + l_3 = 0$, and the j's must satisfy the triangular inequalities

$$|j_1 - j_2| \le j_3 \le j_1 + j_2, \qquad \text{etc.}$$

Then the wave function of the system is

$$\psi_0 = \sum_{l_1, l_2, l_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{pmatrix} \psi_{j_1,l_1} \otimes \psi_{j_2,l_2} \otimes \psi_{j_3,l_3}.$$

Note that the 3-j symbols are zero unless $l_1 + l_2 + l_3 = 0$.

Intensities and Selection Rules

If a hydrogen atom emits a light photon by transition from one energy level to another, the frequency ν is determined by the rule

$$\hbar \nu = E_n - E_m = \frac{\alpha^2 \varepsilon}{2} \left(\frac{1}{m^2} - \frac{1}{n^2} \right), \tag{7.77}$$

provided $n > m$; see (7.54, 55). The observed spectra of hydrogen are known as the *Balmer series* when $m = 1$, which is in the ultraviolet part of the electromagnetic spectrum, the *Lyman series* when $m = 2$, which is in the visible part, the *Paschen series* when $m = 3$, which is in the infrared part. The theory and data fit very accurately. The same relation holds when the atom absorbs energy from a photon of frequency ν.

The intensities of these spectral lines are measured by a standard quantum mechanics formula. For dipole radiation, the probability of an atom jumping from a state ψ to a

state χ with lower energy in a given time period is proportional to the sum of the norms of the "matrix elements" of the position operator $\mathbf{Q}$, i.e.,

$$| \langle \psi, \mathbf{Q}\chi \rangle |^2.$$

The intensity of the emitted light corresponding to this transition is proportional to the transition probability. In particular, if the matrix elements are all 0, then there can be no transition. This leads to certain "selection rules" which prescribe which transitions are possible, by saying which matrix elements are non-zero.

Now suppose we are in a system with rotational symmetry. Suppose ψ corresponds to azimuthal quantum number j and χ corresponds to azimuthal quantum number k, i.e., ψ lies in the representation space corresponding to the irreducible representation ρ^j and similarly for χ. Now the operator $\mathbf{Q}$ corresponds to multiplication by the linear polynomial x, whose angular components correspond to the three-dimensional representation ρ^1. Therefore, we can regard $\mathbf{Q}\chi$ as an element of the tensor product representation $\rho^1 \otimes \rho^k$, which decomposes as the direct sum of three representations

$$\rho^1 \otimes \rho^k = \rho^{k+1} \oplus \rho^k \oplus \rho^{k-1} \qquad \text{unless} \qquad k = 0 \ \text{ or } \ \tfrac{1}{2}.$$

In the exceptional cases,

$$\rho^1 \otimes \rho^0 = \rho^1, \qquad \rho^1 \otimes \rho^{1/2} = \rho^{3/2} \oplus \rho^{1/2}.$$

Now, according to the orthogonality conditions, the elements of different representation spaces are always orthogonal. Therefore, the matrix elements $\langle \psi, \mathbf{Q}\chi \rangle$ are necessarily zero unless $j = k+1, k$, or $k-1$. Therefore, only the following transitions of azimuthal quantum number are allowed:

$$j \ \longmapsto \ j-1, \ j, \ j+1, \quad \text{with the transition } \ 0 \longmapsto 0 \ \text{ not allowed.} \tag{7.78}$$

To get the selection rules for the magnetic quantum number, we need to look at how the individual basis elements in the tensor product decompose. According to the construction of the Wigner coefficients (7.71), we see that the basis element $\psi_{j,l}$ will be represented as a linear combination of $\psi_{k,m}$'s in the tensor product representation $\rho^j \otimes \rho^k$ only when $l = m-1, m$, or $m+1$. Therefore we have the corresponding selection rule for the magnetic quantum number

$$l \ \longmapsto \ l-1, \ l, \ l+1. \tag{7.79}$$

This implies that, in the Zeeman effect (7.60) governing the splitting of spectral lines due to a magnetic field, instead of the $(2j+1)(2j'+1)$ lines one would expect without the selection rules, the intensities of these lines are all zero (in the approximation we are using) and so one would only expect to observe three lines in all, symmetrically arranged around the unperturbed lines. However, many of the observations do not follow this rule, but are rather manifestations of the "anamolous Zeeman effect" which requires electron spin. One can, through a more detailed analysis of the Wigner coefficients, find the different probabilities of the various transitions. Also, physically, the emitted photon corresponding to the transition $m \longmapsto m$ is polarized in the z direction, whereas in the other two cases,

an observer in the xy plane will observe linearly polarized light, whereas an observer in the z direction will observe circularly polarized light. This is because the first transition only gives nonzero values for the operator Z corresponding to the z coordinate, while the other two correspond to the operators $X \pm \mathrm{i}\,Y$.

Parity

For most physical systems, the Hamiltonian is not just invariant under the rotation group $\mathrm{SO}(3)$, but under the full orthogonal group $\mathrm{O}(3)$, consisting of rotations and reflections. Topologically, $\mathrm{O}(3)$ consists of two disjoint copies of $\mathrm{SO}(3)$ with the second component generated by multiplying a rotation by the reflection represented by $-\mathrm{I}$. Therefore, it is not hard to see that the representations of $\mathrm{O}(3)$ will be determined by representations of $\mathrm{SO}(3)$ along with a representation of $-\mathrm{I}$, which must take one of the values $\pm\mathrm{I}$ on the representation space.

The complete list of representations of $\mathrm{O}(3)$ is labelled by a non-negative integer $j = 0, 1, 2, \ldots$, and a sign ± 1, where

$$\widehat{\rho}^{\,j,\pm}(A) = \rho^j(A), \qquad \widehat{\rho}^{\,j,\pm}(-A) = \pm\,\rho^j(A), \qquad A \in \mathrm{SO}(3). \qquad (7.80)$$

The sign $\pm$ associated with a state ψ in a particular representation $\widehat{\rho}^{\,j,\pm}$ is known as the *parity* of the state. For systems with $\mathrm{O}(3)$ symmetry, it is a conserved quantity for the Schrödinger equation. As with all non-infinitesimal quantum mechanical objects, it has no classical analogue.

Interestingly, only half of the representations of $\mathrm{O}(3)$ appear among the spherical harmonics, which is why we failed to observe it when we treated the hydrogen atom before. Indeed, the representation corresponding to the spherical harmonics with total angular momentum j consists of polynomials (in x, y, z) of degree j, and so has parity $(-1)^j$. The other representations cannot arise among scalar-valued functions. Thus, in a scalar wave theory, only the representations $\widehat{\rho}^{\,0,+}, \widehat{\rho}^{\,1,-}, \widehat{\rho}^{\,2,+}, \widehat{\rho}^{\,3,-}, \ldots$ actually occur. Under tensor products, the parities multiply, i.e., $\widehat{\rho}^{\,i,\delta} \otimes \widehat{\rho}^{\,j,\varepsilon}$ where $\delta, \varepsilon = \pm 1$, will decompose into the direct sum of $\widehat{\rho}^{\,k,\kappa}$, where k runs from $|\,i - j\,|$ to $i + j$, and $\kappa = \delta\,\varepsilon$. Thus, two particles both in a state of even or of odd parity will combine to give an even parity state, whereas if the particles have opposite parity, the combined system will have odd parity.

This immediately gives the selection rule for parity. Since the position operator $\mathbf{Q}$ corresponds to the representation $\widehat{\rho}^{\,1,-}$ given by the linear polynomials, the tensor product $\widehat{\rho}^{\,1,-} \otimes \widehat{\rho}^{\,j,\varepsilon}$ will decompose into three (or two or one) representations all with opposite parity to $\widehat{\rho}^{\,j,\varepsilon}$. Therefore, the selection rule for parity is

$$\varepsilon \longmapsto -\varepsilon,$$

i.e., for any transition to occur, the system must change parity! In particular, for the scalar wave theory of the hydrogen atom, the selection rule $j \longmapsto j$ is ruled out by reasons of parity. More generally, this is known as *Laporte's Rule*: for $\mathrm{O}(3)$ symmetric systems governed by the scalar wave equation, the azimuthal quantum number can only change by an odd integer.

In terms of the spectral lines of the hydrogen atom, this gives the observed transitions. Recall that each energy level E_n has states labelled s, p, d, f, etc. corresponding to azimuthal quantum numbers $j = 0, 1, \ldots, n-1$. For hydrogen, the energy levels are given by the formula

$$E_n = -\frac{\beta}{n^2}, \tag{7.81}$$

for some constant β. For an alkali atom, e.g., lithium, sodium, potassium or cesium, which has just one valence electron moving in a central field due to the nucleus and the other electrons, it is observed that the energy levels are given by a formula of the form

$$E_{n,j} = -\frac{\beta}{(n-k)^2}, \tag{7.82}$$

where k depends on j (and, technically speaking, n as well, but not significantly). This formula follows from the perturbation theory. The wave lengths of alkali spectral lines can be obtained from terms like these, which were arranged into series, called s, p, d, f series, of the form $1s, 2s, 3s, \ldots, 2p, 3p, 4p, \ldots, 3d, 4d, 5d, \ldots, \ldots$, where the integer indicates the value of the principal quantum number n. Now, the selection rule (7.78) for j says that transitions only occur between neighboring series, i.e., for radiation, we can have a transition from a p state to an s state or an f state, but we cannot have a transition from a p state to another p state. Similarly, an s state can only change into a p state. Note that the main quantum number n is still allowed to change in any fashion (although it must decrease for emission, and increase for radiation, as energy must be conserved). This is what is physically observed.

Spin

Spin manifests itself in quantum mechanics as an "intrinsic angular momentum" of a quantum particle or system. It is needed to explain a number of experiments whose results do not corroborate with the scalar form of the Schrödinger equation. One can, however, account for these observations by assuming that an electron has its own internal angular momentum or spin which does not depend on its orbital motion or azimuthal quantum number. This was originally proposed by Goudsmit and Uhlenbeck in the 1920's. If the direction of magnetization of a ferro-magnetic bar is reversed, the bar acquires an angular momentum. The observed ratio between the induced angular momentum and the magnetic moment are as $\hbar$ to 2μ, where μ is Bohr's magneton (7.61), rather than as $\hbar$ to μ, as it would be if the magnetization were due to the orbital angular momenta of the electrons. To explain this using electron spin, one assumes that the magnetic moment of a spinning electron is twice as large as the magnetic moment of a corresponding state with only angular momentum.

The Stern–Gerlach experiment is another important indicator of the existence of spin. In this experiment, a beam of silver atoms in the ground state is sent into a magnetic field varying in a perpendicular direction. The beam splits into two components, corresponding to the values of $\pm\mu$ for the internal magnetic moment. It can be assumed that only one electron in this ion is responsible for the magnetic moment because the spins of the other electrons cancel out. (This assumption is plausible because a silver ion Ag^+ shows no

Zeeman effect.) This implies that the magnetic moment of the electron in any direction can only take on the discrete values $\pm\mu$, i.e., the angular momentum is quantized. Moreover, since the ratio of magnetic moment to angular momentum is as μ to $\frac{1}{2}\hbar$, it follows that the mechanical spin momentum in any direction is $\pm\frac{1}{2}\hbar$. Now, in the theory developed above, the azimuthal quantum number j for the single electron could assume any integral value, and for such a value there are $2j + 1$ different values of the magnetic quantum number m. The perturbation due to a magnetic field would then split the beam into $2j + 1$ smaller beams. But the splitting into just two beams does not fit this theory, and why no beam corresponding to $\mu = 0$ existed under the perturbation remained unexplained by the spinless theory.

Another experiment which does not fit the spin-free theory is the anomalous Zeeman effect. Alkali atoms — the series with nonzero angular momentum $j \neq 0$, i.e., the p, d, etc. series — are observed to have the following fine structure. The individual spectral line is composed of two very close spectral lines, known as a doublet. Moreover, if the atom is placed into a magnetic field, one of the lines of each doublet splits into $2j + 2$ lines, whereas the other line splits into $2j$ lines. The existence of a doublet could be explained in the spin-free theory by allowing two electrons in each distinct energy level. However, each electron would have azimuthal quantum number j and so each line in the doublet would split into $2j + 1$ finer lines under the external magnetic field.

All of these anomalous effects can be easily explained by introducing the concept of electron spin. Essentially, spin is an intrinsic angular momentum associated with any quantum mechanical particle or system. In contrast to the classical angular momentum, spin is allowed to assume both integral and half-integral values in the quantum theory, so the underlying symmetry group is SU(2), the double covering of SO(3). The different values of spin correspond to the different irreducible representations of SU(2). Because the spin is proportional to $\hbar$, it goes to 0 as $\hbar \longrightarrow 0$, and so has no classical analogue.

In order to incorporate spin into a quantum system, one postulates that the wave function ψ of the system depends not only on the spatial coordinates $\mathbf{q} = (x, y, z)$, but also on a discrete "spin variable" σ, so we write

$$\psi(\mathbf{q}, \sigma) \qquad \text{or} \qquad \psi_\sigma(\mathbf{q}),$$

for the wave function; alternatively, we can regard the ψ_σ as a vector containing the spin components of the wave function, indexed by the spin variable σ. Their spatial integral

$$\int |\psi_\sigma(\mathbf{q})|^2 \, d\mathbf{q},$$

determines the probability that the particle has a certain value σ of the spin. Similarly, the probability that the particle is in a given region $\Omega \subset \mathbb{R}^3$ is given by the integral

$$\int_\Omega \sum_\sigma |\psi_\sigma(\mathbf{q})|^2 \, d\mathbf{q},$$

found by summing over all possible values of the spin.

The *spin operator* $\mathbf{S} = (S_x, S_y, S_z)$, when applied to the wave function, acts on the spin variables. Since it satisfies the same commutation relations as the angular momentum

operator $\mathbf{J}$, i.e.,

$$[S_x, S_y] = \mathrm{i}\, S_z, \qquad [S_z, S_x] = \mathrm{i}\, S_y, \qquad [S_y, S_z] = \mathrm{i}\, S_x, \qquad (7.83)$$

we can decompose its representation on the state space into irreducible representations. Since our discussion of the representation theory of the Lie algebra $\mathfrak{su}(2)$ only depended on the commutation relations, the same results hold for the spin operator. Each irreducible representation is characterized by the eigenvalues $s(s+1)$ of the square of the spin operator $S^2 = S_x^2 + S_y^2 + S_z^2$, where the *spin quantum number* s can take on any integral or half integral value. Moreover, the irreducible representation ρ^s corresponding to a given value of s, is $(2s+1)$–dimensional, with canonical basis provided by the eigenvectors of the z-component of the spin S_z, whose eigenvalues are $-s, -s+1, \ldots, s-1, s$. Accordingly, the wave function of a particle of a given spin s has $2s+1$ components. Thus a particle of spin 0, e.g., a p-meson, has a single wave function, the spin is well-defined and the theory reduces to the standard Schrödinger theory. The majority of elementary particles have spin $\frac{1}{2}$, including electrons, protons, neutrons, μ-mesons, etc., and so are represented by a wave function with two spin components: $\psi_{1/2}(\mathbf{x}) = (\psi_{1/2,1/2}(\mathbf{x}), \psi_{1/2,-1/2}(\mathbf{x}))$, indexed by the two possible values $\pm\frac{1}{2}$ of the z-component. (The x and y components also take on the possible values $\pm\frac{1}{2}$, but their eigenvectors are certain linear combinations of the components of ψ. In particular, if the particle has a well defined spin in one direction, the spin has equal probability of being $\pm\frac{1}{2}$ in the other two coordinate directions.) Spin 1 particles, e.g., photons, are characterized by three wave functions, $\psi = (\psi_{1,-1}, \psi_{1,0}, \psi_{1,1})$ and so on for higher spins corresponding to other particles or nuclei. In general, the components of the spin operator act on the components $\psi_{s,t}$, $t = -s, -s+1, \ldots, s-1, s$, of the wave function as our earlier angular momentum operators did, cf. (7.26):

$$
\begin{aligned}
S^2 \psi_{s,t} &= s(s+1)\,\psi_{s,t}, & S_+ \psi_{s,t} &= \sqrt{s(s+1) - t(t+1)}\;\psi_{s,t+1}, \\
S_z \psi_{s,t} &= t\,\psi_{s,t}, & S_- \psi_{s,t} &= \sqrt{s(s+1) - t(t-1)}\;\psi_{s,\sigma-1}.
\end{aligned}
\qquad (7.84)
$$

Here

$$S_+ = S_x + \mathrm{i}\, S_y, \qquad S_- = S_x - \mathrm{i}\, S_y. \qquad (7.85)$$

In the case of a particle with spin $\frac{1}{2}$, the action of the spin operator $\mathbf{S}$ coincides with $\frac{1}{2}$ multiplication by the corresponding Pauli matrices (7.32).

For a particle with spin, its total angular momentum, denoted by $\mathbf{L}$, is the sum of the orbital angular momentum operator $\mathbf{J}$ and the spin operator $\mathbf{S}$:

$$\mathbf{L} = \mathbf{J} + \mathbf{S}. \qquad (7.86)$$

Thus $\mathbf{L}$ is also a general "angular momentum operator", and acts on the wave functions according to the tensor product rule for the angular momentum operators $\mathbf{J}$ and $\mathbf{S}$. Therefore, for a particle with orbital angular momentum j and spin s, the total angular momentum can take on the values $l = |\,j - s\,|, |\,j - s\,| + 1, \ldots, j + s$. In the case of an electron with orbital angular momentum j, the total angular momentum can be $l = j \pm \frac{1}{2}$, unless $j = 0$, in which case $l = \frac{1}{2}$ is the only possible value. Similar results hold for systems of particles; one sums the individual spin operators to get the total spin, and the individual orbital

angular momentum operators to get the total orbital angular momentum operator for the system; the sum of these latter two operators is the total angular momentum for the system. Note that a system consisting of an even number of particles always has integral spin, and hence integral total angular momentum.

Thus, the group SU(2) will act on the space of "spinor fields" $\psi = (\psi_\sigma)$ with infinitesimal generators $\mathrm{i}\,\mathbf{L} = \mathrm{i}\,(\mathbf{J} + \mathbf{S})$, given by the tensor product of the standard three-dimensional physical representation ρ^1 on the angular momentum component $\mathbf{J}$, and the spin representation ρ^s on the spin components $\mathbf{S}$. Its action on the physical coordinates reduces to the standard rotations SO(3), whereas its action on the spinor components remains SU(2) for half-integral spin, and reduces to SO(3) in the case of integral spin. In other words, the representation is

$$\rho(A)\,\psi(\mathbf{q}) = \rho^s(A)\,\psi(R^{-1}\,\mathbf{q}), \qquad \text{where} \qquad R = \pi(A) \in \mathrm{SO}(3).$$

For example, if we have a particle of spin $\frac{1}{2}$, and we "rotate" the coordinate system through an angle 2π, i.e., $A = -\mathrm{I}$, then the physical variables remain fixed, while the spin components of the wave function change sign. However, a rotation through 4π returns everything to its original state. The same holds for any particle with half-integral spin; in the case of integral spin, everything is unchanged under just 2π rotations.

Remark: In fact, this behavior is even manifested as a classical phenomenon, known as the "spinor spanner". Attach an object, e.g., a spanner or, as it is known in North America, a wrench, by strings to the walls of a container. If the object is rotated about an axis through an angle of 2π, one cannot disentangle the strings without moving the object, whereas a rotation through 4π does allow a disentangling! (Try it.) This is, of course, another manifestation of the double connectivity of the rotation group SO(3).

The claim is that physical rotations of a system will act on a particle with spin according to the exponential map of the given representation of the total angular momentum operator $\mathbf{L}$. For instance, consider a particle of spin $s = \frac{1}{2}$, which has wave function $\psi(\mathbf{q}) = \big(\psi_-(\mathbf{q}), \psi_+(\mathbf{q})\big)$, where the $\pm$ subscripts are shorthand for the indices $\sigma = \pm\frac{1}{2}$, respectively, and we suppress s. Under a rotation of our coordinate system $\widetilde{\mathbf{q}} = R\,\mathbf{q}$, the wave function will transform into the rotated wave function $\widetilde{\psi}(\widetilde{\mathbf{q}}) = A\,\psi(R^{-1}\widetilde{\mathbf{q}}) = A\,\psi(\mathbf{q})$, where $A = \rho(R)$ is a linear transformation (2×2 matrix) depending on the rotation R. Now the quantity

$$\big(\,|\,\psi_-(\mathbf{q})\,|^2 + |\,\psi_+(\mathbf{q})\,|^2\,\big)\,d\mathbf{q}$$

measures the probability of the particle being found in the infinitesimal volume $d\mathbf{q}$. Since this must be invariant under rotations of the coordinate system, we conclude that

$$|\,\psi_-(\mathbf{q})\,|^2 + |\,\psi_+(\mathbf{q})\,|^2 = |\,\widetilde{\psi}_-(\widetilde{\mathbf{q}})\,|^2 + |\,\widetilde{\psi}_+(\widetilde{\mathbf{q}})\,|^2.$$

This implies that the matrix $A = \rho(R)$ is unitary. Note also that there is some indeterminacy in the definition of A since each state is only determined up to a multiplication of a phase factor $e^{\mathrm{i}\alpha}$, hence we conclude that $\rho(R)$ forms a "projective unitary representation" of the orthogonal group. Moreover, we can assume that $\rho(R)$ has determinant 1 by multiplying it by a suitable phase factor. Composing two rotations, we conclude that

$$\rho(R\,S) = \pm\,\rho(R)\,\rho(S),$$

where the $\pm$ follows because both sides must have determinant 1, so the only freedom is to replace A by $-A$. Thus, ρ determines a (possibly) double-valued two-dimensional representation of the rotation group, hence it is either trivial, or the representation $\rho^{1/2}$. But we know from experiment that a rotation around the z-axis through an angle π reverses the spin, so the action cannot be trivial. Therefore, ρ is the representation generated by the spin operator $\mathbf{S}$.

The invariance of the probability density under rotations motivates a general definition of an algebra of spinors analogous to the standard tensor algebra of covariant and contravariant tensors. Note the analogy of the equation governing the action of spinor fields under elements of SU(2) with the behavior of vector fields $\mathbf{v}$ under ordinary rotations:

$$\rho(R)\,\mathbf{v}(\mathbf{q}) = R\,\mathbf{v}(R^{-1}\,\mathbf{q}), \qquad \text{where} \qquad R \in \mathrm{SO}(3).$$

Since the representations of SU(2) are given by polynomials in the basic representation $\rho^{1/2}$, we deduce that if we multiply spinor fields together, we get "decomposable" higher spin fields. Thus, given a spinor $\psi = (\psi_-, \psi_+)$ corresponding to a linear polynomial in u, v, its tensor square $\psi^{\otimes 2} = (\psi_-^2, 2\,\psi_-\psi_+, \psi_+^2)$ will transform according to the representation ρ^1, once we rewrite it in terms of the orthonormal basis $u^2/\sqrt{2}$, uv, $v^2/\sqrt{2}$. In other words, the square of a spinor field of spin $\frac{1}{2}$ is an ordinary vector field. In this way we can index spinor fields of arbitrary spin s by the powers of the basic spin $\frac{1}{2}$ fields.

Consider an electron in a state with a fixed angular momentum j, corresponding to the representation ρ^j of SO(3), where j is integral. These correspond to the $2\,j + 1$ solutions of the scalar Schrödinger equation. Now, for a spinning electron, if we neglect the interaction of spins and orbits, its components $\psi_\pm$ are each solutions to the same scalar Schrödinger equation, hence they each assume the $2\,j + 1$ possible values. Therefore, the full wave function ψ transforms according to the tensor product representation $\rho^{1/2} \otimes \rho^j$. This decomposes into $\rho^{j+1/2} \oplus \rho^{j-1/2}$, unless $j = 0$, in which case there is only one term, $\rho^{1/2}$. The states with $j = 0$ are known as singlet states, whereas those with $j > 0$ are known as doublets, owing to this splitting. This is manifested in the fine structure of the spectral lines of the alkali atoms, where the lines split into two for the doublet states, but remain one for singlet states. For atoms with several (free) electrons, there are further possible splittings, triplet, etc., corresponding to combinations of these electrons into states of higher spin. Moreover, if the atom is placed into a magnetic field, one of the lines of each doublet splits into $2\,j + 2$ lines, corresponding to the representation $\rho^{j+1/2}$, whereas the other line splits into $2\,j$ lines, corresponding to the representation $\rho^{j-1/2}$. Note that the spin-free theory would predict that each line corresponding to a given value of j would split into $2\,j + 1$ finer lines under a magnetic field. Thus, the spin theory explains the anomalous Zeeman effect.

Proceeding as we did earlier analyzing the normal Zeeman effect, recall that the spin-free quantum Hamiltonian had the form

$$\mathcal{H} = \mathcal{H}_0 + \mu\,\mathbf{H} \cdot \mathbf{J}.$$

Here $\mathcal{H}_0$ denotes the Hamiltonian for the free atom, μ is the Bohr magneton, $\mathbf{H} = h_0\,\mathbf{k}$ is a uniform field, which is taken parallel to the z-axis, and $\mathbf{J}$ denotes the usual "orbital"

angular momentum operator, and we are (as before) ignoring a small term. To incorporate spin, we use the modified operator

$$\mathcal{H} = \mathcal{H}_0 + \mu\,\mathbf{H}\cdot(\mathbf{J} + 2\,\mathbf{S}) = \mathcal{H}_0 + \mu\,\mathbf{H}\cdot(\mathbf{L} + \mathbf{S}),$$

where $\mathbf{S}$ is the spin operator. The factor of 2 follows from the observed correspondence between the spin and the mechanical angular momentum, or can be derived directly from the relativistic theory. The extra term involving the spin operator $\mathbf{S}$, which now couples the Schrödinger equations for two components of the wave function ψ is referred to as spin-orbit coupling, which refers to the fact that the spin and the magnetic fields interact. Note that we have now moved beyond our earlier "naïve quantization", in which we could readily replace classical Hamiltonians by simple quantum ones according to the standard quantum rules. The spin terms have no classical analogue, and so cannot be derived by inspection of the classical Hamiltonian.

The change in energy ΔE due to the splitting is determined as follows: recall that we only needed to determine the eigenvalues of the modified Hamiltonian, which was originally

$$\mathcal{H} = \mathcal{H}_0 + \mu\,h_0\,J_z,$$

but in the spin theory has become

$$\mathcal{H} = \mathcal{H}_0 + \mu\,h_0\,(J_z + 2\,S_z).$$

On the irreducible $(2\,j + 1)$–dimensional space corresponding to the angular momentum j, the operators $\mathbf{J}, \mathbf{S}$ commute, and so determine two irreducible representations of $\mathfrak{su}(2)$. According to Schur's lemma, they are equivalent representations, i.e., there is an invertible map taking one to the other. However, since the operators J_z and S_z commute, they can be simultaneously diagonalized, and so, by Schur's lemma, the representations of $\mathfrak{su}(2)$ corresponding to the operators $\mathbf{J}$ and $\mathbf{S}$ must be proportional:

$$\mathbf{J} = \lambda\,\mathbf{S}.$$

In particular, if we label the states by the representation corresponding to $\mathbf{L}$, then

$$\mathbf{J} + 2\,\mathbf{S} = g\,\mathbf{L},$$

where the number g is called the *Landé factor*. The energy levels corresponding to the different magnetic quantum numbers will then be

$$E_{n,j} - g\,\mu\,h_0\,l = -\frac{m\,e^4}{2\,\hbar\,n^2} - g\,\mu\,h_0\,l, \qquad l = -j,\ \ldots\ ,j, \tag{7.87}$$

differing from the spin-free theory (7.60) only by the Landé factor.

To compute the Landé factor g, we note the elementary identity

$$\mathbf{L}\cdot(\mathbf{J} + 2\,\mathbf{S}) = \tfrac{3}{2}L^2 + \tfrac{1}{2}S^2 - \tfrac{1}{2}J^2,$$

since $\mathbf{L} = \mathbf{J} + \mathbf{S}$, hence, on the given representation space

$$\mathbf{L}\cdot(\mathbf{J} + 2\,\mathbf{S}) = \tfrac{3}{2}\,l\,(l+1) + \tfrac{1}{2}\,s\,(s+1) - \tfrac{1}{2}\,j\,(j+1).$$

On the other hand,

$$\mathbf{L} \cdot (\mathbf{J} + 2\,\mathbf{S}) = g\,L^2 = g\,l\,(l+1).$$

Therefore the Landé factor is

$$g = \frac{3}{2} + \frac{s\,(s+1) - j\,(j+1)}{2\,l\,(l+1)}. \tag{7.88}$$

Note that if there is no spin, so $s = 0$ and $j = l$, then $g = 1$. On the other hand, if the orbital angular momentum is 0, then $g = 2$. We have

$$\Delta E = \frac{m\,g\,\mu}{h_0}.$$

Note that if $g = 0$, then splitting does not occur.

This result, coupled with the earlier selection and intensity rules serves to describe the observed spectral lines of the alkali elements — except for the "hyperfine structure" due to interactions with the nucleus.

Systems of Identical Particles

When a quantum mechanical system consists of two or more indistinguishable particles, some special considerations come into play. In classical mechanics, identical particles do not lose their individuality, despite the fact that they have identical physical properties. Once we label each individual particle, we can follow it through any interaction, and say with precision which particle is which at all times. In quantum mechanics, this does not apply, since we cannot measure everything about a particle at any instant. Thus, when we observe two electrons at one time, and so label them, and then we make another observation at a later time, there is no way to say which electron is which. Thus in quantum mechanics, identical particles have completely lost their individuality, and there is no way of distinguishing them with certainty.

Consider first a system of two particles. If the particles are different, then the wave function of the system will be the product $\psi(\xi_1, \xi_2) = \psi_1(\xi_1)\,\psi_2(\xi_2)$, where $\psi_1(\xi)$ represents the individual state of the first particle and $\psi_2(\xi)$ the state of the second particle. Here ξ is used to denote the spatial coordinates (x, y, z), as well as any spin coordinates that might be required. The underlying state space is just the tensor product of the individual state spaces. Now consider what happens if the particles are identical. Suppose we make an observation of the system, and determine that one of the particles is in the state given by the wave function $\psi_1(\xi)$ while the other particle in the state given by the wave function $\psi_2(\xi)$. Since the particles are identical, there is no way of determining which of the two particles is in the first state, and which in the second. In practice, these wave functions will be localized in different regions of space, so that the combined wave functions

$$\psi(\xi_1, \xi_2) = \psi_1(\xi_1)\,\psi_2(\xi_2), \qquad \widetilde{\psi}(\xi_1, \xi_2) = \psi_1(\xi_2)\,\psi_2(\xi_1),$$

are independent and represent two possible states of the system giving the same observation. Thus, the observation does not allow us to distinguish between the state ψ and the state $\widetilde{\psi}$, or, for that matter, any linear combination $\alpha\,\psi + \beta\,\widetilde{\psi}$ of these two states. Thus

we are unable to precisely determine the state of the system by observation, and we say that there is an *exchange degeneracy* of the system.

In order to analyze this situation, it is useful to introduce the symmetric and anti-symmetric combinations of these two possible states:

$$\psi^S(\xi_1, \xi_2) = \frac{\psi_1(\xi_1)\,\psi_2(\xi_2) + \psi_2(\xi_1)\,\psi_1(\xi_2)}{\sqrt{2}}\,, \qquad \psi^S(\xi_2, \xi_1) = \psi^S(\xi_1, \xi_2),$$

$$\psi^A(\xi_1, \xi_2) = \frac{\psi_1(\xi_1)\,\psi_2(\xi_2) - \psi_2(\xi_1)\,\psi_1(\xi_2)}{\sqrt{2}}\,, \qquad \psi^A(\xi_2, \xi_1) = -\,\psi^A(\xi_1, \xi_2),$$

the factors $\sqrt{2}$ coming from the normalization constraint. Note that since the Hamiltonian must be invariant under the interchange of particles, if the system starts out in a symmetric or anti-symmetric state, it must remain there. In general, the system will be in a linear combination of these two states:

$$y = \alpha\,\psi^S + \beta\,\psi^A, \qquad |\,\alpha\,|^2 + |\,\beta\,|^2 = 1,$$

where the parameters α, β will be constant in time. The probability density of finding one particle at position ξ_1 and the other at position ξ_2 is

$$|\,\psi(\xi_1, \xi_2)\,|^2 + |\,\psi(\xi_2, \xi_1)\,|^2 = 2\left(|\,\alpha\,|^2\,|\,\psi^S(\xi_1, \xi_2)\,|^2 + |\,\beta\,|^2\,|\,\psi^A(\xi_1, \xi_2)\,|^2\right).$$

Now, if the probabilities are going to be independent of the particular linear combination of symmetric and anti-symmetric wave functions giving the state of the system, this quantity must be independent of the parameters α, β. This only happens if

$$|\,\psi^S(\xi_1, \xi_2)\,| = |\,\psi^A(\xi_1, \xi_2)\,|.$$

However, the latter equation is rarely true. For instance, if the two wave packets overlap, then at the point $\xi_1 = \xi_2$ we have $\psi^S(\xi_1, \xi_1) \neq 0$, whereas $\psi^A(\xi_1, \xi_1) = 0$, which is not consistent with the above equation. Therefore, the existence of an exchange degeneracy is a real problem, as it does not allow one to make predictions of the statistical distribution of measurements of the system, even after observing the initial state of the two particles.

One way to overcome this difficulty is to introduce a *symmetrization postulate* that, for a system of two particles, the possible states are necessarily either symmetric or anti-symmetric. Particles that are symmetrical under interchange are said to obey *Bose–Einstein statistics*, and are called *bosons*, whereas those that are anti-symmetrical under interchange are said to obey *Fermi–Dirac statistics*, and are called *fermions*. (The statistics refer to the statistical mechanics of a large number of these particles. Ordinary classical particles obey yet another kind of statistics, in which one can distinguish the particles by label, called *Maxwell–Gibbs statistics*.) According to relativistic quantum mechanics, particles with an integral spin are always bosons, whereas those with half integral spin are always fermions. Note that an important consequence of the nature of fermions is the *Pauli Exclusion Principle*: two fermions cannot be in the same state. This is because any anti-symmetric wave function must vanish identically when $\xi_1 = \xi_2$. Note that as a consequence, we deduce that in any system, two electrons cannot have the same quantum numbers n, j, l, and the same spins. This leads to the explanation of the periodic table.

For a system of two particles the symmetric and anti-symmetric states are the only two possibilities, since we can always decompose any wave function of two particles into a purely symmetrical and a purely anti-symmetrical part. For three or more particles, things are even more complicated, as this is no longer true. Indeed, the number of possible permuted versions of the product $\psi_1(\xi_1)\,\psi_2(\xi_2)\,\cdots\,\psi_n(\xi_n)$ is $n!$, but there is only one completely symmetrical linear combination, and only one completely anti-symmetrical linear combination. The other possible combinations decompose into other representations of the group of permutations, as discussed below.

Let S_n denote the *symmetric group* consisting of all $n!$ permutations of $\{1, 2, \ldots, n\}$. The Hamiltonian must be invariant under permutations of the particles, i.e., $\mathcal{H} \cdot \widehat{\pi} = \widehat{\pi} \cdot \mathcal{H}$ for every $\pi \in S_n$, where

$$\widehat{\pi}\,\psi(\xi_1, \ldots, \xi_n) = \psi(\xi_{\pi 1}, \xi_{\pi 2}, \ldots, \xi_{\pi n}). \tag{7.89}$$

Therefore, we can decompose the space of solutions to the associated Schrödinger equation into irreducible representation spaces for the symmetric group S_n. It is thus of great interest to study the representation theory of this group. The *Symmetrization Hypothesis* states that only the completely symmetric or completely anti-symmetric wave functions are of interest. Thus in systems of identical particles, many of the possible eigenvalues of the Hamiltonian or energies of the system are excluded on symmetrization grounds.

One way to physically justify this restriction to purely symmetrized or anti-symmetrized wave functions is as follows. The normalized wave function of a system of two particles has the form $\psi(\xi_1, \xi_2)$. Interchanging the particles amounts to interchanging their coordinates, giving the wave function $\psi(\xi_2, \xi_1)$, which is also clearly normalized. Now the principle of indistinguishability of identical particles says that $\psi(\xi_1, \xi_2)$ and $\psi(\xi_2, \xi_1)$ must be the same physical state, hence

$$\psi(\xi_1, \xi_2) = \alpha\,\psi(\xi_2, \xi_1),$$

where α is some complex number of modulus 1. Performing the interchange again, we deduce that $\alpha^2 = 1$, hence $\alpha = \pm 1$. In the case $\alpha = 1$, the wave function $\psi(\xi_1, \xi_2)$ is symmetrical with respect to interchange, whereas in the case $\alpha = -1$, the wave function $\psi(\xi_1, \xi_2)$ is anti-symmetrical with respect to interchange. Similar arguments hold for systems of more than two particles. The wave function $\psi(\xi_1, \ldots, \xi_n)$ of a system of n bosons must be symmetrical under permutations, and is so given by

$$\psi(\xi_1, \ldots, \xi_n) = \frac{1}{\sqrt{n!}} \sum_{\pi \in S_n} \psi_1(\xi_{\pi 1})\,\psi_2(\xi_{\pi 2})\,\cdots\,\psi_n(\xi_{\pi n}). \tag{7.90}$$

Similarly, the wave function of n fermions takes the form

$$\psi(\xi_1, \ldots, \xi_n) = \frac{1}{\sqrt{n!}} \sum_{\pi \in S_n} (\mathrm{sign}\,\pi)\,\psi_1(\xi_{\pi 1})\,\psi_2(\xi_{\pi 2})\,\cdots\,\psi_n(\xi_{\pi n}) = \frac{1}{\sqrt{n!}} \det\big[\,\psi_i(\xi_j)\,\big], \tag{7.91}$$

the latter expression known as a *Slater determinant*. Again, the Pauli exclusion principle holds: no two fermions can be in the same state, as otherwise the combined wave function would vanish.

The analysis of systems of identical particles in more detail requires a knowledge of the representation theory of the symmetric group S_n. Intimately tied to this theory is the theory of symmetry classes of tensors, of which the completely symmetric and completely anti-symmetric ones form just a small part. These will now be discussed, but first we need to present a brief outline of the representation theory of finite groups, which also has applications to the crystallographic groups, molecules, etc.

Representation Theory of Finite Groups

References: [**15**], [**19**], [**37**], [**53**], [**54**].

Given a finite group G with $n = \#G$ elements, the *group ring* R is identified with the n-dimensional complex vector space $\mathbb{C}^n$. Its basis elements can be labelled by the group elements, so we conventionally write

$$x = \sum_{g \in G} x_g \, g$$

for an element of R, where the $x_g \in \mathbb{C}$ are complex numbers. Thus, there is a one-to-one correspondence between elements of the group ring and complex-valued functions $x : G \longrightarrow \mathbb{C}$, with $x(g) = x_g$. Note that G acts on R by right multiplication

$$\rho^*(h) \cdot x = \sum_{g \in G} x_g(g \cdot h^{-1}), \quad \text{or, equivalently,} \quad \left(\rho^*(h) \cdot x \right)(g) = x(g \cdot h) \quad \text{for} \quad h \in G.$$

This defines a representation of G, called the (right) *regular representation* ρ^*. (There is also a left regular representation.) Since G is finite, ρ^* can be made into a unitary representation on R using the inner product

$$\langle \, x \, , y \, \rangle = \frac{1}{n} \sum_{g \in G} x(g) \, \overline{y(g)}, \qquad x, y \in R, \tag{7.92}$$

which also makes the left regular representation unitary.

As with any representation of a finite group, the right regular representation decomposes into a direct sum of irreducible representations. The claim is that every irreducible representation of G appears in this sum.

Theorem 7.18. *Let ρ^λ acting on V^λ, for $\lambda = 1, \ldots, m$, be a complete system of inequivalent irreducible representations of a finite group G. Then $m < \infty$. Moreover, the right regular representation ρ^* decomposes*

$$\rho^* = \bigoplus_\lambda n_\lambda \rho^\lambda, \qquad n_\lambda = \dim V^\lambda, \tag{7.93}$$

where the representation ρ^λ appears in ρ^* with multiplicity n_λ equal to its dimension. Thus,

$$\sum_\lambda n_\lambda^2 = n = \#G. \tag{7.94}$$

Proof: Each irreducible representation ρ^λ can be taken to be unitary by suitably introducing an inner product on V^λ. Relative to orthonormal bases of the representation spaces, the matrix elements of the representations $\rho_{ij}^\lambda(g)$, for $i, j = 1, \ldots, n_\lambda$, form an orthogonal system of functions on G, as was shown above for compact groups. (The same proof goes through with averaging over the group replacing the invariant integral.) Therefore, each function ρ_{ij}^λ can be viewed as an independent element of the group ring R, i.e.,

$$\rho_{ij}^\lambda \simeq \sum_{g \in G} \rho_{ij}^\lambda(g) \cdot g.$$

and, as such, they are all linearly independent. Therefore, they are bounded in number by the dimension of R. Since each ρ^λ has n_λ^2 distinct matrix elements, we find

$$\sum_\lambda n_\lambda^2 \leq n = \dim R = \#G.$$

In particular, this proves that there are at most finitely many irreducible representations. We next show that the $\rho_{ij}^\lambda(g)$ form a basis for R. Indeed, let $M \subseteq R$ be the subspace spanned by the matrix elements. The subspace M is invariant under the right (and left) regular representation of G, since

$$\rho^*(h) \cdot \rho_{ij}^\lambda = \sum_{g \in G} \rho_{ij}^\lambda(g) \, g \cdot h^{-1} = \sum_{g \in G} \rho_{ij}^\lambda(g \cdot h) \, g = \sum_{g \in G} \sum_k \rho_{ik}^\lambda(g) \rho_{kj}^\lambda(h) \cdot g = \sum_k \rho_{kj}^\lambda(h) \, \rho_{ik}^\lambda,$$

so $\rho^*(h) \cdot \rho_{ij}^\lambda$ is a linear combination of the matrix elements and so lies in M. Note further that, for each $1 \leq i \leq n_\lambda$, the elements $v_1 = \rho_{i1}^\lambda, \ldots, v_{n_\lambda} = \rho_{in_\lambda}^\lambda$ of M can be regarded as the basis vectors for a copy of the representation ρ^λ, since by the definition of matrix elements,

$$\rho^\lambda(h) \, v_j = \sum_k \rho_{kj}^\lambda(h) \, v_k.$$

Therefore, there are n_λ copies of each irreducible representation ρ^λ contained in M, and so M has the direct sum decomposition into irreducible representations given for R in the statement of the theorem.

It remains to show that $M = R$. Consider the orthogonal complement to M (relative to the inner product making the right regular representation unitary). Since M is invariant, $M^\perp$ is therefore also invariant, and therefore decomposes into irreducible representations. We can therefore find a suitable orthonormal basis $v_1, \ldots, v_{n_\lambda} \in M^\perp$ for one of the irreducible representations ρ^λ contained in $M^\perp$, transforming under G according to the associated matrix elements. However, since $g = e \cdot g$,

$$v_j(g) = [\rho^*(g) v_j](e) = \sum_k \rho_{kj}^\lambda(h) \, v_k(e),$$

hence

$$v_j = \sum_k v_k(e) \, \rho_{kj}^\lambda,$$

which is a linear combination of the matrix elements, and hence is an element of M. This is a contradiction unless $M^\perp = \{0\}$, which completes the proof. *Q.E.D.*

Our earlier results concerning the characters of the irreducible representations also hold here. In particular, the characters form an orthonormal subset of functions on G, i.e., an orthonormal basis of a subspace of R. It is easy to characterize this subspace.

Theorem 7.19. *The characters of the irreducible representations form an orthonormal basis for the subspace*

$$S = \left\{\, x \in R \,\middle|\, x(h^{-1}gh) = x(g) \,\right\}$$

consisting of functions on G which are invariant under conjugation.

Proof: Let $x \in S$. Expand x in terms of matrix elements

$$x = \sum_{i,j,\lambda} a_{ij}^\lambda \rho_{ij}^\lambda.$$

Then, since x is conjugation-invariant,

$$x(g) = \frac{1}{n}\sum_{h \in G} x(h^{-1}gh) = \frac{1}{n}\sum_{h \in G}\sum_{i,j,\lambda} a_{ij}^\lambda \, \rho_{ij}^\lambda(h^{-1}gh)$$

$$= \frac{1}{n}\sum_{h \in G}\sum_{i,j,k,l,\lambda} a_{ij}^\lambda \, \rho_{ik}^\lambda(h^{-1}) \, \rho_{kl}^\lambda(g)\, \rho_{lj}^\lambda(h) = \sum_{i,j,k,l,\lambda} a_{ij}^\lambda \, \rho_{kl}^\lambda(g)\frac{1}{n}\sum_{h \in G}\overline{\rho_{ki}^\lambda(h)}\,\rho_{lj}^\lambda(h)$$

$$= \sum_{i,j,k,l,\lambda} a_{ij}^\lambda \, \rho_{kl}^\lambda(g) \, \langle\, \rho_{ki}^\lambda\,,\rho_{lj}^\lambda\,\rangle = \sum_{i,k,\lambda}\frac{a_{ii}^\lambda}{n_\lambda}\rho_{kk}^\lambda(g) = \sum_{i,\lambda}\frac{a_{ii}^\lambda}{n_\lambda}\chi^\lambda(g).$$

Thus x is a linear combination of the irreducible characters.   *Q.E.D.*

Corollary 7.20. *The number of inequivalent irreducible representations of a finite group G equals the number of conjugacy classes in G.*

Definition 7.21. Let ρ be a representation of G on the vector space V. An equivariant *projection operator* is a map $P\colon V \longrightarrow V$ satisfying $P^2 = P$, and $P \cdot \rho(g) = \rho(g) \cdot P$ for all $g \in G$.

The range and kernel of a projection operator are easily seen to be invariant subspaces for the representation ρ, with $V = \operatorname{rng} P \oplus \ker P$, and so ρ is irreducible if and only if there are no non-trivial projection operators. ($P = \mathrm{I}$ and $P = 0$ are trivial projection operators.) Furthermore, $\operatorname{rng} P$ is an irreducible representation of G if and only if P cannot be decomposed into the sum of two projection operators $P = P_1 + P_2$ satifying $P_1 P_2 = 0 = P_2 P_1$ and $P_1, P_2 \neq 0$.

In the special case of the group ring under the right regular representation, there is another characterization of projection operators. An element of the group ring $\varepsilon \in R$ is called an *idempotent* if it satisfies $\varepsilon \cdot \varepsilon = \varepsilon$.

Lemma 7.22. *Every projection operator for the right regular representation of G on its group ring R is given by left multiplication by the idempotent $\varepsilon = Pe$.*

Proof: Since P commutes with right multiplication,

$$P\,x = \sum x_g P\,g = \sum x_g(P\,e)\,g = \sum x_g\,\varepsilon\,g = \varepsilon\,x. \qquad\qquad Q.E.D.$$

Note that the image $V = \varepsilon R = \{\,\varepsilon\,x \mid x \in R\,\}$ of an idempotent is invariant under the right regular representation, and hence is a representation of G. Note also that e itself is an element of V, being the image of e. The idempotent is *primitive* if it cannot be expressed as the sum of two nonzero idempotents, $\varepsilon = \varepsilon_1 + \varepsilon_2$, with $\varepsilon_1^2 = \varepsilon_1$, $\varepsilon_2^2 = \varepsilon_2$, $\varepsilon_1\varepsilon_2 = \varepsilon_2\varepsilon_1 = 0$. Primitive idempotents determine irreducible representations and conversely. It is useful to have an alternative characterization of a primitive idempotent.

Lemma 7.23. . *An element $\varepsilon \in R$ is a primitive idempotent if and only if $\varepsilon\,x\,\varepsilon = k\varepsilon$ for every $x \in R$, where $k = k_x \in \mathbb{C}$ is a scalar (depending on x).*

Proof: Let $V = \operatorname{rng} P_\varepsilon = \{\,x\varepsilon \mid x \in R\,\}$. Then the operator $y \longmapsto y\varepsilon x\varepsilon$ commutes with the left action of g and maps V to itself. Therefore, by Schur's lemma, when restricted to V, it must be a multiple of the identity map and hence is a multiple of P_ε on all of R. Conversely, if we have such an idempotent $\varepsilon = \varepsilon_1 + \varepsilon_2$, with $\varepsilon_1\varepsilon_2 = \varepsilon_2\varepsilon_1 = 0$, then $\varepsilon\varepsilon_1\varepsilon = (\varepsilon_1 + \varepsilon_2)\varepsilon_1(\varepsilon_1 + \varepsilon_2) = \varepsilon_1$, hence ε_1 is a multiple of ε, which implies that either ε_1 or ε_2 is zero, so ε is primitive. $\qquad\qquad Q.E.D.$

Lemma 7.24. *Two irreducible representation subspaces V and $\widetilde{V}$ define equivalent representations if and only if their primitive idempotents have the property that there exist non-zero elements of the form $\widetilde{\varepsilon}\,x\varepsilon \neq 0$ for some $x \in R$.*

Proof: Let $A\colon V \xrightarrow{\;\sim\;} \widetilde{V}$ be an equivalence. Let $y = A(\varepsilon) \in \widetilde{V}$. Then for any $v \in V$,

$$A(v) = A(\varepsilon\,v) = A(\varepsilon)\,v = y\,v,$$

so the equivalence is given by left multiplication by $y \neq 0$. Moreover $\widetilde{\varepsilon}\,y\varepsilon = y$, so y is of the proper form. Conversely, given $y = \widetilde{\varepsilon}\,x\varepsilon \neq 0$, it is easy to see that left multiplication by y commutes with the right regular representation, and so, by irreducibility, defines an equivalence between the two subspaces. $\qquad\qquad Q.E.D.$

Finally, an element σ is called an *essential idempotent* if $\sigma^2 = k\,\sigma$ for some nonzero scalar $k \neq 0$. Note that in that case, $\varepsilon = \sigma/k$ is then an idempotent. Therefore, the classification of irreducible representations of a finite group can be reduced to the determination of a sufficient number of primitive essential idempotents of the group ring.

Representations of the Symmetric Group

The *symmetric group* S_n consists of the permutations of $\{\,1, 2, \ldots, n\,\}$. A *transposition* is the permutation $(i\,j)$ which interchanges i and j. Any permutation can be decomposed into a product of transpositions; if $\pi \in \mathrm{S}_n$ can be written as a product of k transpositions, its sign is defined as $\operatorname{sign} \pi = (-1)^k$. (The decomposition and the number k are not unique, but the sign is well-defined.) A *cycle* is a permutation $(i\,j\,k\ldots l\,m)$ which takes i to j, j to k, $\ldots$, l to m and m to i. Any permutation can be uniquely written as the product of disjoint cycles. For example, the permutation p taking $(1\,2\,3\,4\,5\,6)$ to $(5\,1\,6\,4\,2\,3)$ can

be written as $p = (125)(36)(4)$. (The cycles of length 1, e.g., (4), are just the identity element.) Since disjoint cycles commute, the order is irrelevant, and we assume that they are arranged in decreasing order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. The *cycle length* of a permutation is the non-increasing sequence of integers $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, giving the lengths of the cycles. Note that $\lambda_1 + \lambda_2 + \cdots + \lambda_m = n$, so λ is a *partition* of n. It is easy to show that two permutations π and $\widetilde{\pi}$ are conjugate, i.e., $\pi = s\,\widetilde{\pi}\,s^{-1}$, if and only if they have the same cycle length. Therefore, the conjugacy classes of S_n are indexed by the partitions λ of n. Therefore, to each partition of n there corresponds a unique conjugacy class, and hence a unique irreducible representation of the symmetric group S_n.

It is convenient to represent a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, $\lambda_1 + \lambda_2 + \cdots + \lambda_m = n$, by a corresponding *shape* or *Young diagram*[†], which consists of a a planar arrangement of n boxes, with λ_1 boxes in the first row, λ_2 boxes in the second row, etc. For each shape λ, we construct a tableau by filling in the boxes in λ by the integers $1, 2, \ldots, n$ in any order. Clearly there are $n!$ tableaux corresponding to each shape. A tableau is called standard if its rows and columns are strictly increasing. The number n_λ of standard tableaux of shape λ will play a key role, and is given by the *hook length formula* due to Frame, Robinson, and Thrall, [**23**]:

$$n_\lambda = \frac{n!}{\displaystyle\prod_{i,j \in \lambda} h_{ij}}. \tag{7.95}$$

The denominator is the product of the hook lengths for all boxes (i,j) in λ, where the $(i,j)^{\text{th}}$ hook consists of all the boxes (i,j), (i',j), $i' > i$, (i,j'), $j' > j$ which lie in λ, and the hook length is the number of such boxes. This formula can be straightforwardly proved by induction on n.

Given a tableau T let $R(T) \subset S_n$ denote the permutations which just permute the rows of T and $C(T) \subset S_n$ denote the permutations which just permute the columns of T. Define the element

$$\sigma_T = \sum_{\tau \in R(T)} \sum_{\chi \in C(T)} (\text{sign}\,\chi)\,\tau\,\chi, \tag{7.96}$$

known as the *Young symmetrizer* associated with the tableau T.

Theorem 7.25. *For each tableau T, the element σ_T is a primitive essential idempotent. The invariant subspace $V_T = \sigma_T R$ is equivalent to the irreducible representation space ρ^λ for S_n corresponding to the shape λ. The dimension n_λ of ρ^λ, and its multiplicity in the right regular representation ρ^*, equals the number of standard tableau of shape λ:*

$$\rho^* = \bigoplus_\lambda n_\lambda\,\rho^\lambda, \qquad \text{and so} \qquad R = \bigoplus_{T \text{ standard}} V_T. \tag{7.97}$$

[†] Named after the early twentieth century mathematician Alfred Young, and not an indication of its youth!

Proof: Given a tableau T and $\pi \in S_n$, let πT denote the tableau obtained by permuting the entries of T according to π. It is not hard to show that

$$R(\pi T) = \pi R(T) \pi^{-1}, \qquad C(\pi T) = \pi C(T) \pi^{-1}, \qquad \text{hence} \qquad \sigma_{\pi T} = \pi \sigma_T \pi^{-1}.$$

This immediately implies that if T and $\widetilde{T}$ have the same shape λ, and hence there is a permutation π taking T to $\widetilde{T}$, then the representations corresponding to V_T and $V_{\widetilde{T}}$ are isomorphic. $\hspace{5cm}$ *Q.E.D.*

Next we investigate the structure of the components of the Young idempotent.

Lemma 7.26. *Let T be a tableau. A permutation $\pi \in S_n$ can be written as $\pi = \tau \chi$ where $\tau \in R(T)$, $\chi \in C(T)$, if and only if no two entries which lie in the same row of T end up in the same column of πT.*

Proof: First suppose $\pi = \tau \chi$. An important tricky point throughout the proof is that $\tau \chi T$ is not obtained by first using χ to permute the columns of T, and then using τ to permute the rows of χT, since τ is not necessarily in $R(\chi T)$. (Try this on an example!) However, we can rewrite $\tau \chi T = (\tau \chi \tau^{-1}) \tau T$, and first use τ to permute the rows of T, and then, by the above remark, $\tau \chi \tau^{-1} \in C(\tau T)$ will permute the columns of τT. In particular, since τT has the same rows as T, but arranged in different order, and $\tau \chi \tau^{-1}$ acts on τT by permuting its columns, the different row elements of T must end up in different columns of $\pi T = \tau \chi T = (\tau \chi \tau^{-1}) \tau T$.

Conversely, if no two elements of the same row of T lie in the same column of πT, the integers in the first column of πT lie in different rows of T, so we can perform a row permutation on T to put them all into the first column. Similarly for the second and subsequent columns of πT. Therefore, there is a permutation $\tau \in R(T)$ such that τT has the same column elements as πT, but in different order. This implies that there is a column permutation $\widetilde{\chi} \in C(\tau T)$ such that $\pi T = \widetilde{\chi} \tau T$. but $C(\tau T) = \tau C(T) \tau^{-1}$. Thus, $\widetilde{\chi} = \tau \chi \tau^{-1}$ for some $\chi \in C(T)$, hence $\pi = \widetilde{\chi} \tau = \tau \chi \tau^{-1} \tau = \tau \chi$, as desired. $\hspace{2cm}$ *Q.E.D.*

Lemma 7.27. *An element $x \in R$ satisfies*

$$\tau x \chi = (\operatorname{sign} \chi) x, \qquad \text{for all} \qquad \tau \in R(T), \qquad \chi \in C(T), \tag{7.98}$$

if and only if $x = k \sigma_T$ for some scalar k.

Proof: First note that σ_T satisfies (7.98) because

$$\tau \sigma_T = \sigma_T, \qquad \tau \in R(T), \qquad \sigma_T \chi = (\operatorname{sign} \chi) \sigma_T, \qquad \chi \in C(T).$$

Conversely, note that (7.98) implies that

$$x(\tau \chi) = (\operatorname{sign} \chi) x(e), \qquad \tau \in R(T), \qquad \chi \in C(T).$$

Setting $k = x(e)$, we need only prove that $x(\pi) = 0$ whenever π is not of the form $\tau \chi$. But this implies that there are two entries in some row of T which lie in the same column of πT. Let τ denote the transposition of these two entries, so $\tau \in R(T)$, and $\chi = \pi^{-1} \tau \pi \in C(T)$ is also a transposition. Then, using (7.98), $x(\pi) = x(\tau \pi \chi) = - x(\pi)$. $\hspace{1.5cm}$ *Q.E.D.*

As a corollary, we prove that the Young symmetrizer σ_T is a primitive essential idempotent. Indeed, from our earlier equations

$$\tau \, \sigma_T^2 \chi = \tau \, \sigma_T \, \sigma_T \chi = (\operatorname{sign} \chi) \, \sigma_T^2, \qquad \tau \in R(T), \qquad \chi \in C(T),$$

so by the lemma, $\sigma_T^2 = k \, \sigma_T$. To show $k \neq 0$, it suffices to prove that the trace of the linear transformation defined by σ_T is not 0:

$$\operatorname{tr} \sigma_T = \sum_\pi \, \langle \, \pi \, , \sigma_T \, \pi \, \rangle = \sum_\pi \, 1 = n!$$

Finally, to show the primitiveness, we only need show $y = \sigma_T x \sigma_T$ is a multiple of σ_T for every $x \in R$. But

$$\tau \, y \, \chi = \tau \, \sigma_T \, x \, \sigma_T \, \chi = (\operatorname{sign} \chi) \, \sigma_T \, x \, \sigma_T = (\operatorname{sign} \chi) \, y, \qquad \tau \in R(T), \qquad \chi \in C(T),$$

so by the lemma y is a multiple of σ_T.

The last item to check is that the irreducible representations corresponding to tableau having different shapes are not equivalent. By Lemma 7.24, it suffices to prove that if T and $\widetilde{T}$ are tableaux with different shapes, then $\sigma_{\widetilde{T}} x \sigma_T = 0$ for all $x \in R$. First note that if T has two entries in the same row which lie in the same column of $\widetilde{T}$, then $\sigma_{\widetilde{T}} \sigma_T = 0$, since the transposition of these two entries, denoted $\tau \in R(T) \cap C(\widetilde{T})$, which implies that $\sigma_{\widetilde{T}} \sigma_T = \sigma_{\widetilde{T}} \tau \tau \sigma_T = -\sigma_{\widetilde{T}} \sigma_T$. Now, if these tableaux have different shapes, then it is not hard to see that, by possibly interchanging the two tableau, this condition will always hold. (Just look at the first row of different length, and assume that that row of T is longer than that of $\widetilde{T}$.) The same holds for πT, so $\sigma_{\widetilde{T}} \pi \sigma_T = \sigma_{\widetilde{T}} \sigma_{\pi T} \pi = 0$ for all permutations π, which is enough to prove inequivalence. *Note*: We can also reprove equivalence of the representations corresponding to tableau of the same shape. Choose $\pi \in \mathrm{S}_n$ mapping T to $\widetilde{T}$. Then $\sigma_{\widetilde{T}} \pi \sigma_T = \pi \sigma_T \pi^{-1} \pi \sigma_T = k \pi \sigma_T \neq 0$. $\hfill$ Q.E.D.

Symmetry Classes of Tensors

References: [4], [13], [15], [33], [34], [53], [54], [56], [61].

There is a deep connection between the representation theory of the symmetric group and the classification of symmetry classes of tensors, based on the representation theory of the general linear group. Note first that if W is a finite-dimensional vector space, the general linear group $\mathrm{GL}(W)$ consisting of all invertible linear transformations $A \colon W \longrightarrow W$ is *not* a compact Lie group, and hence our earlier statements about representations for finite groups and compact Lie groups do not necessarily apply. Indeed, it is not true that every representation of $\mathrm{GL}(W)$ is unitary; nor is it true that all irreducible representation are finite-dimensional; nor is it true that every reducible representation decomposes as the direct sum of two representations! However, there is an important special class of representations of $\mathrm{GL}(W)$ for which analogues of these statements hold, the "polynomial representations" or "tensor representations".

Let

$$\otimes^n W = W \otimes \, \cdots \, \otimes W$$

denote the n-fold tensor product of W with itself. Then $\mathrm{GL}(W)$ has a natural representation on this space, namely the *tensor product representation*, which is uniquely determined by its action on decomposable tensors:

$$(\otimes^n A)(w_1 \otimes \cdots \otimes w_n) = (A\,w_1) \otimes \cdots \otimes (A\,w_n), \qquad A \in \mathrm{GL}(W). \tag{7.99}$$

As we will see, the tensor product space $\otimes^n W$ decomposes into a direct sum of irreducible subspaces, and, moreover, this decomposition is (essentially) identical with the decomposition of the regular representation of the symmetric group S_n.

Indeed, S_n acts on $\otimes^n W$ by permuting the factors:

$$\rho_W(\pi)(w_1 \otimes \cdots \otimes w_n) = w_{\pi^{-1}1} \otimes \cdots \otimes w_{\pi^{-1}n}, \qquad \pi \in \mathrm{S}_n. \tag{7.100}$$

(The inverse is needed to make ρ_W a true representation.) We can therefore decompose $\otimes^n W$ into irreducible representations of S_n:

$$\rho_W = \bigoplus_\lambda k_\lambda \rho^\lambda,$$

where the direct sum is over all partitions λ of the integer n.

The group ring of S_n acts on $\otimes^n W$ by linearity. In particular, the Young symmetrizers (7.96) will determine projection operators

$$\sigma_T \colon \otimes^n W \longrightarrow \otimes^n W.$$

The range of σ_T will be denoted by W_T. Since the actions of S_n and $\mathrm{GL}(W)$ commute, W_T is invariant under $\mathrm{GL}(W)$, and so forms a representation space. The claim is that this is an irreducible representation. Moreover, two such subspaces are equivalent representations if and only if their tableaux have the same shape. Since the subspaces will be independent if and only if the corresponding projections are independent, we conclude that there is a one-to-one correspondence between standard tableaux and the different irreducible components of the tensor product space.

Theorem 7.28. *The tensor product representation of $\mathrm{GL}(W)$ decomposes into a direct sum of irreducible representations*

$$\otimes^n W = \bigoplus_\lambda n_\lambda L_\lambda W \tag{7.101}$$

where the sum is over all shapes which have n boxes and at most $\dim W$ rows. The representation space $L_\lambda W$ corresponding to the shape λ has multiplicity n_λ equal to the number of standard tableaux of shape λ. Projection operators realizing this decomposition are provided by the Young symmetrizers corresponding to all the standard tableaux.

In particular, when $\lambda = (n)$ has only one row, then $L_\lambda W = \odot^n W$ is the nth *symmetric power* of W, that is, the space of fully symmetric tensors. On the other hand, when $\lambda = (1, 1, \ldots, 1)$ has only one column, then $L_\lambda W = \wedge^n W$ is the nth *exterior power* of W, that is, the space of fully anti-symmetric tensors. Each of these occurs with multiplicity $n_\lambda = 1$ in (7.101) because there is only one standard tableau of the given shape, namely that obtained by filling in the integers $1, \ldots, n$ in increasing order along the row or column.

In fact, one can use the fact that $\otimes^n W$ is a representation space for the product group $\mathrm{S}_n \times \mathrm{GL}(W)$ to give a more invariant description of the above decomposition:

$$\otimes^n W = \bigoplus_\lambda M_\lambda \otimes L_\lambda W \tag{7.102}$$

where M_λ is the representation space, known as the *Specht module*, corresponding to the irreducible representation ρ^λ of S_n. In this form, S_n only acts on the M_λ's, whereas $\mathrm{GL}(W)$ only acts on $L_\lambda W$, the *Weyl module* corresponding to the partition λ. The operation L_λ, which in simple cases gives the symmetrization or skew-symmetrization operation, is known as the *shape functor* or *Schur functor* and has very general functorial properties.

The proof of (7.102) follows from some more general results about commutator algebras. Let ρ be a representation of a finite group G on the vector space V. Then ρ induces a representation of the group ring R on V by linearity:

$$\rho(x)\,v = \sum x_g\,\rho(g)\,v, \qquad v \in V, \qquad x = \sum x_g\,g \in R. \tag{7.103}$$

Let

$$K = K(R) = \{\, A \in \mathcal{L}(V) \mid [A, \rho(x)] = 0 \ \text{ for all } \ x \in R \,\} \tag{7.104}$$

denote the *commutator algebra* of R, which forms an algebra of linear maps of V.

For the action of S_n on $\otimes^n W$, the commutator algebra K is spanned by the tensor product action $\otimes^n \mathrm{GL}(W)$ of the general linear group, i.e.,

$$K = \left\{\, \sum_{i=1}^n c_i\, A_i \otimes \ \cdots \ \otimes A_i \ \middle| \ c_i \in \mathbb{C}, \ \ A_i \in \mathrm{GL}(W) \,\right\}. \tag{7.105}$$

To see this, given a basis $e_1, \ldots, e_m$ of W, we introduce a basis

$$e_I = e_{i_1} \otimes \ \cdots \ \otimes e_{i_n}, \qquad 1 \le i_\nu \le m,$$

of $\otimes^n W$. A linear transformation $B \colon \otimes^n W \longrightarrow \otimes^n W$ is written in matrix form (B_J^I) relative to this basis. It is easy to see that B commutes with the action of S_n on $\otimes^n W$ if and only if its matrix elements are symmetric:

$$B_{\pi(J)}^{\pi(I)} = B_J^I, \qquad \pi \in \mathrm{S}_n.$$

Such transformations are referred to as *bisymmetric*. But this means that B can be viewed as an element of the space of symmetric tensors in the tensor space $\otimes^n \mathcal{L}(W)$ of the vector space $\mathcal{L}(W)$ of linear transformations on W, and it is well known that the set of powers $A \otimes \ \cdots \ \otimes A$ span the space of symmetric tensors over any vector space.

Now, we can decompose V into irreducible representations of G (and hence irreducible representations of R), or into irreducible representations of the commutator algebra K. (The definitions for K are similar.) Remarkably, these two decompositions are, although not identical, essentially "isomorphic", in the following sense.

Theorem 7.29. *Suppose ρ decomposes V into irreducible representation spaces*

$$V = \bigoplus_\lambda k_\lambda V_\lambda,$$

where V_λ is the representation space for the irreducible representation ρ^λ. Then there is a corresponding decomposition of V into irreducible representation spaces for the commutator algebra K:

$$V = \bigoplus_\lambda l_\lambda X_\lambda,$$

where each X_λ is an irreducible representation space for K. The multiplicity k_λ of V_λ is the same as the dimension of X_λ, and the multiplicity l_λ of X_λ is the same as the dimension of V_λ. Moreover, each copy of X_λ can be realized as the image of a primitive idempotent ε_λ in the group ring R, i.e.,

$$X_\lambda = \rho(\varepsilon_\lambda)V = \{\, \rho(\varepsilon_\lambda)v \mid v \in V \,\},$$

corresponding to all independent primitive idempotents $\varepsilon \in R$ such that $\rho(\varepsilon) \neq 0$.

Suppose we decompose

$$V = \bigoplus_\lambda n_\lambda V_\lambda = \bigoplus_\lambda \bigoplus_{i=1}^{n_\lambda} V_\lambda^i,$$

where $\rho \mid V_\lambda^i$ is isomorphic to ρ^λ. Given $B \in K$, we decompose it into block matrix form $(B_{\lambda,\mu}^{i,k})$ relative to this decomposition, where each block

$$B_{\lambda,\mu}^{i,k} \colon V_\mu^k \longrightarrow V_\lambda^i$$

satisfies

$$B_{\lambda,\mu}^{i,k} \rho^\mu(g) = \rho^\lambda(g)\, B_{\lambda,\mu}^{i,k} \qquad \text{for all} \qquad g \in G.$$

By Schur's Lemma 7.5, this implies that $B_{\lambda,\mu}^{i,k} = 0$ if $\lambda \neq \mu$, while $B_{\lambda,\mu}^{i,k}$ is a multiple of the identity if $\lambda = \mu$. Now choose a basis $\{\, e_\lambda^{i,k} \mid k = 1, \ldots, n_\lambda \,\}$ for each V_λ^i. Set

$$X_\lambda^k = \text{span}\,\{ e_\lambda^{i,k} \}, \qquad i = 1, \ldots, n_\lambda.$$

Then each X_λ^k is invariant under K and $S \mid X_\lambda^k = \mathcal{L}(X_\lambda^k)$ consists of all linear transformations, so X_λ^k is an irreducible representation space for K. Note that this immediately implies that the commutator algebra of K coincides with the space of linear transformations restricting to $\mathcal{L}(V_\lambda^i)$ on each irreducible representation space of G. This is the same as the action of the group ring R on V_λ^i since the matrix elements of each irreducible representation ρ^λ form an orthogonal basis of the space $\mathcal{L}(V_\lambda)$, and hence $R \mid V_\lambda^i = \mathcal{L}(V_\lambda^i)$. Thus we have shown that $\rho(R)$ and K are commutators of each other: $K(K(\rho(R))) = \rho(R)$.

Now, consider a $K^\perp$-invariant subspace $U \subset V$. We know that V is the direct sum of suitable V_λ^{k}'s, hence there is an associated invariant complement $U^\perp$, so that $V = U \oplus U^\perp$. Let $P \colon V \longrightarrow V$ be the projection corresponding to this direct sum decomposition of V. Then the invariance of U and $U^\perp$ shows that P commutes with all operators in K, i.e.,

P lies in the commutator algebra of K, so that P corresponds to an idempotent element $e \in R$, with $P = \rho(\varepsilon)$. Moreover, the image of P is an irreducible representation space for K if and only if (a) ε is a primitive idempotent and, (b) $\rho(\varepsilon) \neq 0$. The second condition depends on the precise nature of the representation ρ under consideration.

Lemma 7.30. *Let* $\dim W = m$. *Let* ε *denote the standard representation of* S_n *on* $\otimes^n W$. *Let* ε_T *denote the Young idempotent corresponding to the tableau* T *of shape* λ. *Then* $\rho(\varepsilon_T) = 0$ *if and only if* λ *has more than* m *rows.*

Proof: By suitably relabeling the copies of W in $\otimes^n W$, we can assume without loss of generality that T is the simplest standard tableau of the given shape λ, obtained by filling in λ with the entries 1 to n row by row, so the first row has the integers $1, \ldots, \lambda_1$, in order, the second has $\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2$, etc. Let $e_1, \ldots, e_n$ be a basis for W. If λ has less than m rows, then it is easy to see that the element $e^* = e_1 \otimes \cdots \otimes e_1 \otimes e_2 \otimes \cdots \otimes e_2 \otimes e_3 \otimes \cdots$, having λ_1 copies of e_1 followed by λ_2 copies of e_2, etc., gets mapped to a nonzero element upon multiplication by ε_T. Indeed, applying the first factor $\sum (\operatorname{sign} \chi) \chi$ of ε_T to e^* leads to a combination of basis elements, only one of which, namely e^* itself, still has e_1 in the first λ_1 copies of W in $\otimes^n W$. The second factor $\sum \tau$ only permutes the factors corresponding to the different rows of λ, so the element e^* gets mapped to a nonzero multiple of itself, and none of the other elements in $\sum (\operatorname{sign} \chi) \chi(e^*)$ can cancel this out. On the other hand, since the factor $\sum (\operatorname{sign} \chi) \chi$ skew symmetrizes the columns, if W has fewer basis elements than the number of rows of T, each basis element will be automatically mapped to zero by this factor, and so $\rho(\varepsilon_T) = 0$.   *Q.E.D.*

Tensor Products

The *Clebsch–Gordan series* for the symmetric group S_n will determine the decomposition of the tensor product of two irreducible representations:

$$\rho^\lambda \otimes \rho^\mu = \bigoplus_\nu a_\nu^{\lambda,\mu} \rho^\nu, \tag{7.106}$$

where λ, μ, ν denote shapes corresponding to n. (In combinatorics, this is also known as the "inner product" of the representations ρ^λ, ρ^μ.) Remarkably, very little is known in general about the coefficients $a_\nu^{\lambda,\mu}$, aside from special isolated results, tables for lower order ones, and some computational rules for finding them.

More interesting is the theory of the tensor products of irreducible polynomial representations of the general linear group:

$$L_\lambda W \otimes L_\mu W = \bigoplus_\nu b_{\lambda,\mu}^\nu L_\nu W. \tag{7.107}$$

Note that since $L_\lambda W \subset \otimes^n W$, $L_\mu W \subset \otimes^m W$, where n, m are the number of boxes in λ, μ respectively, then terms in the sum on the right hand side only include the $L_\nu W$ when ν has $n + m$ boxes. Quantum mechanically, the above decomposition corresponds to determining the possible permutational symmetries of a total system when the symmetry classes of two constituent subsystems are known. This decomposition corresponds to an

operation on the representations of the symmetric groups known as the "outer product" to distinguish it from the inner or tensor product. In group theoretic language, the outer product of the irreducible representation ρ^λ of S_n and the irreducible representation ρ^μ of S_m will give a (usually reducible) representation of S_{n+m} obtained by "inducing" the representation of $S_n \times S_m$ on the tensor product $\rho^\lambda \otimes \rho^\mu$ to $S_{n+m} \supset S_n \times S_m$.

The coefficients in the tensor product deomposition (7.107) can be explicitly computed using the *Littlewood–Richardson Rule*. To implement the algorithm, we do the following: First, fill in the (simpler) shape μ with 1's in the first row, 2's in the second, etc. Then append the resulting indexed boxes one at a time to the blank shape λ in order subject to the restrictions:

(i) The resulting diagram is standard (strictly increasing columns; non-decreasing rows.)

(ii) When the symbols are read in order right to left, top to bottom, at no stage does the total number of j's exceed the total number of i's for any $i < j$.

The final collection of shapes constructed this way gives the decomposition of the tensor product. Interestingly, except for the shapes with more than $\dim W$ columns giving zero Weyl modules, this decomposition is independent of the underlying dimension of W.

Example 7.31.
$$W \otimes L_\lambda W = \bigoplus_\mu L_\mu W, \tag{7.108}$$

the sum being over all shapes μ obtained from λ by appending a single box. More generally,

$$\odot^k W \otimes L_\lambda W = \bigoplus_\mu L_\mu W, \tag{7.109}$$

the sum being over all shapes μ obtained from λ by adding on k boxes such that no two of these boxes appear in the same column. These special cases (7.108, 109) are known as the *Pieri formulae*.

A more complicated example is:

$$[21] \otimes [21] = [42] \oplus [41^2] \oplus [3^2] \oplus 2\,[321] \oplus [31^3] \oplus [2^3],$$

whose justification is left as an exercise for the reader.

Suppose we have two vector spaces W, Z. We seek a decomposition of the Weyl modules associated with their tensor product:

$$L_\lambda(W \otimes Z) = \bigoplus_{\mu,\nu} c_\lambda^{\mu,\nu} L_\mu W \otimes L_\nu Z. \tag{7.110}$$

where we are decomposing with respect to $\mathrm{GL}(W) \otimes \mathrm{GL}(Z) \subset \mathrm{GL}(W \otimes Z)$. This formula can be reduced to the earlier question about the tensor product of representations of S_n. We prove that

$$c_\lambda^{\mu,\nu} = a_\nu^{\lambda,\mu},$$

where the a's are as in (7.106). Indeed, we know that

$$\otimes^n(W \otimes Z) = \bigoplus_\lambda V_\lambda \otimes L_\lambda(W \otimes Z).$$

On the other hand,

$$\otimes^n(W \otimes Z) = \otimes^n W \otimes \otimes^n Z = \bigoplus_{\mu,\nu} V_\mu \otimes V_\nu \otimes L_\mu W \otimes L_\nu Z,$$

from which the result follows immediately. The most important cases can be done directly:

$$\wedge^n(W \otimes Z) = \bigoplus_\lambda L_\lambda W \otimes L_{\widetilde{\lambda}} Z, \qquad \odot^n(W \otimes Z) = \bigoplus_\lambda L_\lambda W \otimes L_\lambda Z. \qquad (7.111)$$

In both cases the sum is over all partitions λ of n, and in the first, $\widetilde{\lambda}$ denotes the "dual" partition to λ obtained by interchanging the rows and columns.

Plethysm

Yet another mechanism of combining irreducible representations of the general linear group together is though the operation of *plethysm*. Essentially, given a vector space W, one views $L_\lambda W$ as a vector space in its own right, and so forms $L_\mu(L_\lambda W)$ which is an irreducible representation for $\mathrm{GL}(L_\lambda W)$. Restricting back to $\mathrm{GL}(W)$, this decomposes into irreducible representations:

$$L_\mu(L_\lambda W) = \bigoplus_\nu d^\nu_{\mu,\lambda} L_\nu W. \qquad (7.112)$$

Again, except for special cases, the general form of this plethysm is unknown. However, there are various ways of reducing this formula to just symmetric plethysms. A very useful fact is the formal determinantal formula:

$$L_\lambda W = \text{``}\det\text{''}\left(\odot^{\lambda_i + i - j} W\right). \qquad (7.113)$$

Here, in computing the "determinant" of the indicated $m \times m$ matrix, m being the number of rows in λ, one uses tensor products to multiply the entries, ordered by their row number. For example,

$$L_{(2,1)}W = \text{``}\det\text{''}\begin{pmatrix} \odot^2 W & \odot^3 W \\ \odot^0 W & \odot^1 W \end{pmatrix}$$
$$= \odot^2 W \otimes \odot^1 W - \odot^3 W \otimes \odot^0 W = \odot^2 W \otimes W - \odot^3 W.$$

But by the Pieri formula (7.109), $\odot^2 W \otimes W = L_{(2,1)}W \oplus \odot^3 W$, so the formula is correct. Extensive tables[†] of plethysms can be found in [**12**].

Orbital–Spin and Total Angular Momentum Coupling

These denote different methods of decomposing the states of the electrons of an atom. In the *orbital-spin coupling* scheme (or *LS–coupling*), we want to treat the orbital angular

---

[†] The tables are stated in the combinatorial language of Schur functions, [**33**], but can be directly translated into the identical formulas for Schur functors.

momenta and spins separately. Therefore, we want a decomposition into summands of the form

$$\sigma \otimes \rho^S$$

where the second factor refers to the irreducible representation spaces for the spin operator. The particular summands

$$\rho^J \otimes \rho^S$$

corresponding to a system of electrons with total orbital angular momentum j and total spin s are denoted by symbols of the form ^{2s+1}X, where $X = S, P, D, \ldots$, are the capital versions of the letter series $s, p, d, \ldots$, used to denote the different orbital angular momenta of the individual electrons. The number $2s + 1$ denotes the multiplet for the state, so we get singlet, doublet, triplet, etc. states corresponding to $2s + 1 = 1, 2, 3, \ldots$. Note further that each such state decomposes into irreducible representations

$$\rho^{|j-s|} \oplus \rho^{|j-s|+1} \oplus \cdots \oplus \rho^{j+s},$$

corresponding to the eigenvalues of the total angular momentum operator $\mathbf{L} = \mathbf{J} + \mathbf{S}$. To distinguish these, a subscript is used: $^{2s+1}X_l$. Thus the symbol $^2P_{3/2}$ denotes a system with $j = 1$, $s = \frac{1}{2}$, and $l = \frac{3}{2}$. For example, if we have one electron in the p shell and one in the d shell, this corresponds to the state space

$$\sigma = (\rho^1 \otimes \rho^{1/2}) \otimes (\rho^2 \otimes \rho^{1/2}).$$

If $j \geq s$, then the multiplet will indeed split into $2s + 1$ parts, but for $j < s$, there will be fewer parts than the multiplet terminology indicates. The splitting is basically the same as the splitting of the spectral line under a magnetic field, as in the Zeeman effect for a one electron atom. The initial orbital-spin coupling decomposition leads to

$$\sigma = (\rho^1 \otimes \rho^2) \otimes \rho^0 \oplus (\rho^1 \otimes \rho^2) \otimes \rho^1 = (\rho^1 \oplus \rho^2 \oplus \rho^3) \otimes \rho^0 \oplus (\rho^1 \oplus \rho^2 \oplus \rho^3) \otimes \rho^1$$

corresponding to states $^1P, ^1D, ^1F, ^3P, ^3D, ^3F$. The complete decomposition has states $^1P_1, ^1D_2, ^1F_3$, etc.

Example 7.32. For a helium atom, (or other atom with two free electrons), the orbital angular momentum bound states for a single electron will be indicated by

$$\sigma = \rho^0 \oplus \rho^1 \oplus \rho^2 \oplus \cdots$$

corresponding to the $s, p, d, \ldots$ series. With spin, these become

$$\sigma \otimes \rho^{1/2} = (\rho^0 \otimes \rho^{1/2}) \oplus (\rho^1 \otimes \rho^{1/2}) \oplus (\rho^2 \otimes \rho^{1/2}) \oplus \cdots$$

which can be further decomposed according to the total angular momentum. For two electrons, the Pauli exclusion principle states that we should use the state space

$$\wedge^2(\sigma \otimes \rho^{1/2}) = (\wedge^2 \sigma \otimes \odot^2 \rho^{1/2}) \oplus (\odot^2 \sigma \otimes \wedge^2 \rho^{1/2}),$$

which we decompose according to the principle of orbital-spin coupling. Now

$$\wedge^2 \rho^{1/2} = \rho^0, \qquad \odot^2 \rho^{1/2} = \rho^1,$$

so the state space splits into a sum of two spaces, one having total spin 0, i.e., a singlet, and one having total spin 1, i.e., a triplet. The singlet states can all be regarded as symmetric in the orbital wave functions, while the triplet states are anti-symmetric in the orbital wave functions. Note that the orbital symmetry character χ, corresponding to the two different representations of S_2, will have the selection rule $\chi \longmapsto \chi$, since the position operator $\mathbf{Q}$ amounts to multiplication by the symmetric sum of the individual positions $\mathbf{q} + \widetilde{\mathbf{q}}$ of the two electrons, and so preserves the symmetry of the state. This implies that any matrix elements corresponding to states with different symmetry will be automatically zero. Therefore, spectral lines will only come from transitions within the singlet series and the triplet series, i.e., there cannot be a transition from a triplet term to a singlet term. Similar remarks hold for more complicated electron configurations.

If both electrons have angular momentum j, then we are in the representation space

$$\textstyle\bigwedge^2(\rho^j \otimes \rho^{1/2}) = (\textstyle\bigwedge^2 \rho^j \otimes \bigodot^2 \rho^{1/2}) \oplus (\bigodot^2 \rho^j \otimes \bigwedge^2 \rho^{1/2}) = (\bigwedge^2 \rho^j \otimes \rho^0) \oplus (\bigodot^2 \rho^j \otimes \rho^1).$$

Therefore the singlet terms are anti-symmetric in the orbital wave functions, whereas the triplet terms are symmetric(!). Furthermore, it can be shown that since

$$\textstyle\bigwedge^2 \rho^j \oplus \bigodot^2 \rho^j = \rho^j \otimes \rho^j = \rho^0 \oplus \rho^1 \oplus \ \cdots \ \oplus \rho^{2j},$$

we have the singlet and triplet terms decomposing as

$$\textstyle\bigwedge^2 \rho^j = \rho^1 \oplus \rho^3 \oplus \ \cdots \ \oplus \rho^{2j-1}, \qquad \bigodot^2 \rho^j = \rho^0 \oplus \rho^2 \oplus \ldots \oplus \rho^{2j},$$

respectively. The selection rules for the symmetry character will require that the spectral lines coming from the singlet series and the triplet series, i.e., corresponding to different values of j, are distinct.

Furthermore, we can split the different lines for each series by decomposing the angular momentum pieces of the representation, which can be obtained from the orbital-spin coupling scheme by further decomposing the representations. Alternatively, in the total angular momentum coupling scheme, ("jj coupling") we first decompose $\rho^j \otimes \rho^{1/2}$ according to the usual rules, and then use the known plethysm for the exterior power. For instance, if $j > 0$, then

$$\textstyle\bigwedge^2(\rho^j \otimes \rho^{1/2}) = \bigwedge^2 \big(\rho^{j+1/2} \oplus \rho^{j-1/2} \big) = \bigwedge^2 \rho^{j+1/2} \oplus \big(\rho^{j+1/2} \otimes \rho^{j-1/2} \big) \oplus \bigwedge^2 \rho^{j-1/2}.$$

these can be further decomposed according to the usual rules.

In general, we need to determine a plethysm of the form $L_\lambda \rho^j$, i.e., a "restricted plethysm". There is a trick for $\mathfrak{su}(2) = \mathfrak{so}(3)$ which can reduce this to a nonrestricted plethysm computation. Note that the complexification of this Lie algebra is the same as the complex Lie algebra $\mathfrak{sl}(2) = \mathfrak{sl}(2, \mathbb{C})$, so the representation theory is basically the same. Moreover, the analytic irreducible representations of $\mathrm{GL}(W)$ remain irreducible when restricted to the special linear group $\mathrm{SL}(W)$; the only difference is that some previously different representations become isomorphic. Indeed, it is easy to show that for $\mathrm{GL}(m)$, the representations $[k^m]$ corresponding to a rectangular shape with m rows of length k are all one-dimensional, corresponding to the determinantal representation

$$A \ \longmapsto \ (\det A)^k,$$

and hence, on restriction to $\mathrm{SL}(m)$ are all trivial. Moreover, the Littlewood-Richardson Rule shows that for any shape λ with at most m rows

$$[\lambda + k^m] \simeq [\lambda] \otimes [k^m] \quad \text{as representations of } \mathrm{GL}(m),$$

since all the other terms have more than m rows. Here $\lambda + k^m$ denotes the shape obtained by sticking λ onto the shape k^m, having rows $\lambda_\nu + k$ for $\nu = 1, \dots, m$. Thus, for $\mathrm{SL}(m)$ we can always delete rectangles with m rows from any shape and get an equivalent representation. In the case of $\mathrm{SL}(2)$, this means that all three-rowed shapes reduce to zero, and the two-rowed shape representation (p, q), $p \geq q$, of $\mathrm{GL}(2)$ becomes equivalent to the representation $(p - q, 0) = (p - q)$ when restricted to $\mathrm{SL}(2)$. This observation permits an easy computation of the plethysm formulas for $\mathrm{SL}(2)$, and hence $\mathrm{SO}(3)$, using known plethysm formulas for $\mathrm{GL}(2)$, as given in [**61**].

For example, consider the case of three p electrons. This requires the decomposition of

$$\wedge^3(\rho^1 \otimes \rho^{1/2}) = (\wedge^3 \rho^1 \otimes \odot^3 \rho^{1/2}) \oplus (L_{(2,1)} \rho^1 \otimes L_{(2,1)} \rho^{1/2}) \oplus (\odot^3 \rho^1 \otimes \wedge^3 \rho^{1/2})$$

$$= (\wedge^3 \rho^1 \otimes \rho^{3/2}) \oplus (L_{(2,1)} \rho^1 \otimes \rho^{1/2}).$$

Thus the states split into a doublet, $S = \frac{1}{2}$, and a quadruplet, $S = \frac{3}{2}$. Note that this arises since

$$\otimes^3 \rho^{1/2} = \rho^{1/2} \otimes \rho^{1/2} \otimes \rho^{1/2} = \rho^{1/2} \otimes (\rho^0 \oplus \rho^1) = 2\rho^{1/2} \oplus \rho^{3/2},$$

hence

$$\odot^3 \rho^{1/2} = \rho^{3/2}, \qquad L_{(2,1)} \rho^{1/2} = \rho^{1/2}, \qquad \wedge^3 \rho^{1/2} = 0,$$

the final equality coming from the fact that $\rho^{1/2}$ is only two-dimensional. Alternatively, we can use the identification of $\rho^{1/2}$ with the standard representation of $\mathrm{GL}(2)$ on $W = \mathbb{C}^2$, and so identify $\odot^3 W$ with $\rho^{3/2}$, while $L_{(2,1)} W$ identifieswith W, and hence with $\rho^{1/2}$.

To decompose the orbital angular momentum components, we identify ρ^1 with the representation $\odot^2 W$ of $\mathrm{GL}(2)$, so that we need the plethysm formulae

$$\wedge^3(\odot^2 W) = L_{(3,3)} W \oplus L_{(4,1,1)} W, \qquad L_{(2,1)}(\odot^2 W) = L_{(3,2,1)} W \oplus L_{(4,2)} W \oplus L_{(5,1)} W.$$

Hence, ignoring the three-rowed shapes which all give zero since $\dim W = 2$, and converting the equivalent two-rowed shapes according to the earlier rule, we deduce

$$\wedge^3 \rho^1 = \rho^0, \qquad L_{(2,1)} \rho^1 = \rho^1 \oplus \rho^2.$$

Actually, the first of these is trivial, as ρ^1 is a three-dimensional representation, so $\wedge^3 \rho^1$ is one-dimensional, and hence must be the trivial representation ρ^0. This gives the complete orbital-spin decomposition

$$\wedge^3(\rho^1 \otimes \rho^{1/2}) = (\rho^0 \otimes \rho^{3/2}) \oplus (\rho^1 \otimes \rho^{1/2}) \oplus (\rho^2 \otimes \rho^{1/2}).$$

Therefore, a state describing $3p$ electrons splits into the terms $^4S, ^2P, ^2D$. The sub-terms are then found to be $^4S_{3/2}, ^2P_{1/2}, ^2P_{3/2}, ^2D_{3/2}, ^2D_{5/2}$. Note that the quadruplet term does not actually split any further as it has zero orbital angular momentum, but each of the doublets do split.

In the case of four or more p electrons, there is a trick. Note that if W is a vector space of dimension m, the natural pairing

$$\wedge^k W \otimes \wedge^{m-k} W \; \longrightarrow \; \wedge^m W$$

allows us to identify $\wedge^{m-k}W$ with $\wedge^m W \otimes \wedge^k W^*$, where $*$ denotes the dual space. In particular, if we restrict to $\mathrm{SL}(W)$, then the space $\wedge^m W$ is the trivial representation, hence $\wedge^{m-k}W \simeq \wedge^k W^*$. Finally, if our representation is unitary, then we can identify W with its dual W^*, so $\wedge^{m-k}W \simeq \wedge^k W$. Taking $W = \rho^j \otimes \rho^{1/2}$, corresponding to electrons with orbital angular momentum j, we know that $\dim W = 2(2j+1)$, which is the total number of electrons which can occupy the shell indexed by j. For the p shell, this number is 6. We conclude that the possible states occupied by k electrons having orbital angular momentum j, i.e., the decomposition of $\wedge^k(\rho^j \otimes \rho^{1/2})$, are the same as the states occupied by $2(2j+1)-k$ electrons in the same shell, i.e., the decomposition of $\wedge^{2(2j+1)-k}(\rho^j \otimes \rho^{1/2})$. Thus, the possible states of $4p$ electrons are the same as the states of $2p$ electrons, which we already analyzed. Physicists refer to this as electrons being equivalent to "holes".

8. Relativistic Quantum Mechanics.

References: [**7**], [**8**], [**20**], [**36**].

The fact that velocities are limited by the speed of light introduces new twists in the quantum mechanical interpretation of physical phenomena. In the non-relativistic case, the uncertainty in measuring the energy of a particle or system is related to the length of time in which the measurements are made according to the Uncertainty Relation (4.26):

$$\Delta p \, \Delta q \geq \hbar.$$

For small changes, the speed v is given by

$$v \approx \frac{\Delta q}{\Delta t},$$

and thus

$$v \, \Delta p \; \gtrsim \; \frac{\hbar}{\Delta t}. \tag{8.1}$$

Now if we make a measurement of the momentum of a particle, we must in general affect its velocity. Since the uncertainty in momentum is the same both before and after the measurement, we conclude that the measurement of the momentum of a particle results in a change in its velocity, and the change becomes greater the shorter the duration of the measurement process.

In the relativistic case, the limit imposed by the speed of light implies that we can never measure the momentum of a particle with precision! Now, v is limited by c, and hence, by (8.1),

$$\Delta p \; \gtrsim \; \frac{\hbar}{c \, \Delta t},$$

which gives a limit to the highest accuracy theoretically available in measuring the momentum of a particle in a given time span. In particular, the shorter the time interval in

which we try to measure the momentum, the less accuracy we are able to attain. For a particle whose momentum is changing, this is extremely worrying, since the only way to get a good measurement of the instantaneous momentum is to measure it in a short time period. And an exact measurement is possible only if the measurement takes place over an infinite time interval!?!

In fact, there is as yet no logically consistent and complete relativistic quantum theory. Nevertheless, we are able to apply some basic techniques to analyze the relativistic features of quantum mechanical states, and deduce new features which, remarkably, do compare favorably with experiments. The most striking was Dirac's prediction of the positron, based on his relativistic equation for the electron, and which he initially discounted due to lack of experimental evidence!

The Lorentz Group

The fundamental symmetry groups in special relativity are the linear Lorentz group and the Poincaré group, which includes translations. Here we concentrate on the properties, representation theory, and applications of the former in relativistic quantum mechanical systems.

The pseudo-metric on flat *Minkowski space-time* $\mathbb{R}^4$ is

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2, \tag{8.2}$$

where c is the velocity of light. It is useful to scale time, and introduce the notation $q = (q^\mu)$ for the position vector, with

$$q^0 = ct, \qquad q^1 = x, \qquad q^2 = y, \qquad q^3 = z.$$

Relative to these coordinates, the Minkowski metric has matrix

$$G = (g_{\mu\nu}) = \mathrm{diag}\,(1, -1, -1, -1) = G^{-1} = (g^{\mu\nu}), \tag{8.3}$$

of *signature* $(1,3)$, which indicates the numbers of positive and negative eigenvalues. We define the dual coordinates (using the Einstein summation convention and tensorial raising and lowering of indices throughout)

$$q_\mu = g_{\mu\nu} q^\nu, \qquad \text{so} \qquad q_0 = ct, \qquad q_1 = -x, \qquad q_2 = -y, \qquad q_3 = -z,$$

whereby the Minkowski metric (8.2) is

$$ds^2 = g_{\mu\nu} dq^\mu\, dq^\nu = dq_\mu\, dq^\mu.$$

The linear group $L \subset \mathrm{GL}(4)$ that leaves the Minkowski metric invariant is known as the *Lorentz group*. It consists of the 4×4 matrices A satisfying

$$A^T G A = G. \tag{8.4}$$

In mathematics, the Lorentz group is identified as $L = \mathrm{O}(1,3)$, where, in general, $\mathrm{O}(k,l)$ denotes the linear group of isometries of the diagonal metric of signature (k,l) on $\mathbb{R}^n$, with $n = k + l$. As with the ordinary orthogonal group, $\mathrm{O}(k,l)$ is an $\frac{1}{2}n(n-1)$–dimensional

Lie group; in particular, the Lorentz group is a 6-dimensional Lie group. The Lie algebra of the Lorentz group is the 6-dimensional Lie algebra $\mathfrak{l} = \mathfrak{o}(1,3)$ consisting of all 4×4 matrices X satisfying

$$X^T G + G X = 0. \tag{8.5}$$

For $n \geq 3$, the only other isometries of a nondegenerate quadratic form are the translations. In particular, the group of isometries of the Minkowski metric is known as the *Poincaré group*, and is the semi-direct product of the Lorentz group and the translation group:

$$\mathrm{P}(1,3) = \mathrm{O}(1,3) \ltimes \mathbb{R}^4. \tag{8.6}$$

The defining relation (8.4) for a Lorentz transformation immediately implies that $\det A = \pm 1$. Moreover, if we look at the $(0,0)$ entry, we find

$$a_{00}^2 - a_{10}^2 - a_{20}^2 - a_{30}^2 = 1,$$

hence $|a_{00}| \geq 1$. Thus $L = \mathrm{O}(1,3)$ has four connected components, depending on whether A has determinant ± 1, and whether a_{00} is positive or negative. These are often denoted as

$$L_+^\uparrow, \qquad L_-^\uparrow, \qquad L_+^\downarrow, \qquad L_-^\downarrow,$$

where the $+$ and $-$ refer to the sign of $\det A$, and correspond to spatial orientation, and the $\uparrow$ and $\downarrow$ refer to the sign of a_{00}, and correspond to time orientation. The *proper Lorentz group* $L_+^\uparrow$ is the connected component of the identity. There are three two-component subgroups, all generated by the proper Lorentz group and an additional discrete symmetry. The group $L_+ = \mathrm{SO}(1,3) = L_+^\uparrow \cup L_+^\downarrow$, known as the *special Lorentz group*, requires $\det A = +1$, and is generated by the inversion map $-\mathrm{I}$. The *orthochronous Lorentz group*, $L^\uparrow = L_+^\uparrow \cup L_-^\uparrow$, characterized by $a_{00} \geq 1$, is generated by the *space reversal* or *parity map* P which has matrix G, as in (8.3). The group $L_0 = L_+^\uparrow \cup L_-^\downarrow$, known as the *orthochorous"* *Lorentz group*, is generated by the *time reversal* T which has matrix $-G$. Included in the Lorentz group is the subgroup $\mathrm{O}(3)$ of rotations (and reflections) and the *boosts*, which are hyperbolic rotations between the t-axis and a space axis. Every transformation in the proper Lorentz group can be represented uniquely as BR, where B is a boost and R is a rotation. Note that the boosts are positive definite symmetric matrices, so this result is just a restatement of the polar decomposition of matrices. *Warning*: The boosts do not form a subgroup of L.

We next prove that the Lorentz group is, like the three-dimensional rotation group $\mathrm{SO}(3)$, a doubly connected Lie group. Moreover, its simply connected covering group is the group $\mathrm{SL}(2,\mathbb{C})$ of 2×2 complex matrices of determinant 1, considered as a real Lie group, even though it comes equipped with a complex structure. Note that $\mathrm{SL}(2,\mathbb{C})$ has complex dimension 3, hence real dimension 6, so at least the dimensions match.

Consider the action of the group $\mathrm{SL}(2,\mathbb{C})$ on the space $M(2,\mathbb{C})$ of 2×2 complex matrices given by

$$\rho(A)\, X = A\, X\, A^*, \tag{8.7}$$

where $*$ denotes the Hermitian adjoint (conjugate transpose) matrix. This defines a reducible eight-dimensional representation of $\mathrm{SL}(2,\mathbb{C})$. The subspace of Hermitian matrices

$\mathcal{H} = \{ X \mid X^* = X \}$ forms a four-dimensional invariant subspace of $M(2, \mathbb{C})$. Note that the determinant of X is invariant under the representation ρ. If we coordinatize the subspace $\mathcal{H}$ "correctly", we find

$$\det X = \det \begin{pmatrix} t - z & x + \mathrm{i}\, y \\ x - \mathrm{i}\, y & t + z \end{pmatrix} = t^2 - x^2 - y^2 - z^2, \tag{8.8}$$

which is just the Minkowski length. Therefore (8.7) defines a Lorentz transformation of the matrix X viewed as an element of Minkowski space. Moreover, it is not hard to prove $\rho(A) = \rho(B)$ if and only if $A = \pm B$, so ρ defines a double covering of the Lorentz group by $\mathrm{SL}(2, \mathbb{C})$. Note that on the subgroup $\mathrm{SU}(2) \subset \mathrm{SL}(2, \mathbb{C})$ this action reduces to the earlier adjoint action (7.35) used to construct the double covering of $\mathrm{SO}(3)$.

The discrete transformations generating the full Lorentz group $\mathrm{O}(1, 3)$ have analogues as follows. Note that the transformation corresponding to the matrix

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

produces the Lorentz transformation

$$(t, x, y, z) \ \longmapsto \ (t, -x, y, -z),$$

which is still in the proper Lorentz group. However, the operation

$$C \colon X \ \longmapsto \ \overline{X}$$

of complex conjugation preserves the subspace $\mathcal{H}$, and produces the Lorentz transformation

$$(t, x, y, z) \ \longmapsto \ (t, x, -y, z).$$

Therefore, we can identify $S \circ C$ with the parity reversal Lorentz transformation P. Similarly, the operation

$$\iota \colon X \ \longmapsto \ -X$$

also preserves the subspace $\mathcal{H}$, and produces the Lorentz transformation

$$(t, x, y, z) \ \longmapsto \ (-t, -x, -y, -z),$$

which is the Lorentz transformation $-\mathrm{I}$. Finally, to get time reversal, we use the relation $T = -\mathrm{I}\, P$, so its representative is $\iota \circ S \circ C$. Thus, the group generated by $\mathrm{SL}(2, \mathbb{C})$ combined with the discrete transformations $S \circ C$ and ι forms the simply connected covering group of the full Lorentz group $\mathrm{O}(1, 3)$.

Observe that neither $\mathrm{SL}(2, \mathbb{C})$ nor $\mathrm{O}(1, 3)$ is compact. Thus, not every representation is fully reducible, nor is every representation unitary. For example, the standard representation of the Lorentz group is not unitary as the matrix entries are unbounded. Indeed, except for the trivial representation, no finite-dimensional representation is unitary, but there do exist infinite-dimensional irreducible unitary representations. Doubled-valued representations of the proper Lorentz group appear as ordinary representations of $\mathrm{SL}(2, \mathbb{C})$, and will be discussed later.

The Dirac Equation

We now try to deal with the problem of quantizing the equation for the motion of a relativistic particle. In relativity, the spatial momenta (p_1, p_2, p_3) are the last three components of a four-vector $\mathbf{p} = (p_\mu)$, the *relativistic momentum*, with $p_0 = c^{-1}E$ being the scaled energy. The Minkowski length of this vector is the constant mc, which is positive since the particle moves slower than the speed of light. This leads to the relativistic form of the energy equation

$$p_\mu p^\mu = m^2 c^2, \qquad \text{i.e.,} \qquad E^2 = c^2 p^2 + m^2 c^4,$$

where $p^2 = p_1^2 + p_2^2 + p_3^2$ is the classical momentum. Therefore, the relativistic Hamiltonian is

$$E = + \sqrt{p^2 + m^2 c^2}\,,$$

where we take the positive square root to give the particle positive energy, which indicates that it moves forward in time. In classical relativistic dynamics, a particle with initially positive energy always has positive energy.

To quantize this energy equation, we use the Schrödinger representation

$$P_\mu = -\mathrm{i}\,\hbar \frac{\partial}{\partial q^\mu} \tag{8.9}$$

of the momentum operators, leading to the first order *wave equation*

$$-\mathrm{i}\,\hbar\,\psi_t = \sqrt{-\hbar\,\Delta + m^2 c^2}\,\,\psi, \tag{8.10}$$

where Δ is the usual Laplacian. We will not try to interpret this integro-differential equation too rigorously. As a relativistic model, it has some severe problems, since it is not symmetrical with respect to the four momenta P_μ, so much so that one cannot generalize it in any relativistic way when external fields are present.

An alternative approach is to work with the original energy equation, the quantized wave equation for which then takes the form

$$(P_\mu P^\mu + m^2 c^2)\,\psi = (g^{\mu\nu} P_\mu P_\nu + m^2 c^2)\,\psi = 0. \tag{8.11}$$

Equivalently, we obtain this by applying the pseudo-differential operator

$$-\mathrm{i}\,\hbar\,\partial_t + \sqrt{-\hbar\,\Delta + m^2 c^2}\,. \tag{8.12}$$

to (8.10). In explicit form, the resulting partial differential equation is the well-known *Klein-Gordon equation*

$$-\hbar^2\,\psi_{tt} = -\hbar^2 c^2 \Delta\psi + m^2 c^4\,\psi, \tag{8.13}$$

first proposed as a relativistic quantum equation by Schrödinger. As we will see, the equation is invariant under general Lorentz transformations. It contains the solutions to (8.10), but has additional solutions since the operator (8.12) is not invertible. Moreover, the Klein-Gordon equation now includes the negative energy modes, which are outlawed in special relativity. Classically, this does not present a significant problem due to the

conservation of energy. However, quantum mechanically, this is a serious issue, since the energy can leak from positive to negative values. However, the pricipal difficulty with the Klein-Gordon equation is that it is second order in time, and so the time evolution of a state is not determined solely by its initial value; one must also prescribe $\partial \psi / \partial t$, which does not have a physical interpretation. Moreover, it can only describe particles of spin 0, since there is just one component to the wave function ψ.

Dirac proposed factoring the wave operator, i.e., looking for an equation of the form

$$\sum_{\mu=1}^{4} \gamma^{\mu} P_{\mu} \psi = m c \, \psi, \tag{8.14}$$

whose square is the Klein-Gordon equation (8.13). Multiplying out the factors produces

$$(\gamma^{\mu} P_{\mu} - m c) \, (\gamma^{\nu} P_{\nu} + m c) = \tfrac{1}{2} (\gamma^{\mu} \gamma^{\nu} + \gamma^{\nu} \gamma^{\mu}) \, P_{\mu} P_{\nu} - m^2 c^2.$$

For the summation given by the first term on the right hand side to be the wave operator on Minkowski space, we must have

$$\gamma^{\mu} \gamma^{\nu} + \gamma^{\nu} \gamma^{\mu} = 2 \, g^{\mu\nu}, \tag{8.15}$$

where the $g^{\mu\nu}$ are the components of the Minkowski metric (8.2). The condition (8.15) implies that the γ^{μ} form the generators of the "Clifford algebra" relative to the Minkowski inner product, a concept that was introduced by the English mathematician William Clifford in the mid-nineteenth century. We now turn to a detailed study of Clifford algebras.

Clifford Algebras

References: [**6**], [**16**], [**18**].

Let V be a vector space over a field K (we assume $K = \mathbb{R}$ or $\mathbb{C}$ for simplicity). Let $Q\colon V \to \mathbb{R}$ a quadratic form over V, with associated symmetric bilinear form obtained by polarization:

$$Q(x,y) = \tfrac{1}{2} \big[\, Q(x+y) - Q(x) - Q(y) \,\big], \quad \text{so} \quad Q(x) = Q(x,x), \quad \text{for} \quad x,y \in V. \tag{8.16}$$

Think of $Q(x)$ as a possibly degenerate norm, $\| x \|^2$, with $Q(x,y)$ the associated "inner product" $\langle \, x \,, y \, \rangle$.

Given V and Q, the associated *Clifford algebra* is the quotient space of the tensor algebra

$$\mathcal{C}(Q) = \otimes^{*} V \,/\, \mathcal{I}(Q), \tag{8.17}$$

where $\mathcal{I}(Q) \subset \otimes^{*} V$ denotes the two-sided ideal generated by the relations

$$x \otimes x - Q(x). \tag{8.18}$$

We use juxtaposition to denote the induced product in the Clifford algebra, so $x \, y$ denotes the image of $x \otimes y$ in $\mathcal{C}(Q)$, and $\mathcal{C}(Q)$ is prescribed by the fundamental relation $x^2 = Q(x)$. Note that, by polarizing, this implies the fundamental defining relations

$$x \, y + y \, x = 2 \, Q(x,y). \tag{8.19}$$

Note that if the quadratic form is trivial, $Q(x) \equiv 0$ for all x, then $\mathcal{C}(0)$ is the same as the exterior algebra $\bigwedge^* V$. In all cases, the dimension of the Clifford algebra is the same as the exterior algebra, namely 2^n, where n is the dimension of V, since we can clearly use the products $e_{i_1} e_{i_2} \cdots e_{i_k}$ for $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ as basis elements. Both the scalars K and the vector space V are contained in $\mathcal{C}(Q)$.

From now on, we will always assume that the quadratic form Q is non-degenerate, meaning that $Q(x, y) = 0$ for all $y \in V$ if and only if $x = 0$. For complex vector spaces, there is, up to change of basis, only one non-degenerate quadratic form, so all nondegenerate Clifford algebras of a given dimension are isomorphic. For real vector spaces, the nondegenerate quadratic forms are classified by their signature (k, l), and we denote as $\mathcal{C}(k, l)$ the corresponding real Clifford algebras.

Let us start by investigating what are the possible real, non-degenerate, low-dimensional Clifford algebras. There are two non-degenerate one-dimensional Clifford algebras, denoted

$$\mathcal{C}(1, 0) \simeq \mathbb{R} \oplus \mathbb{R}, \qquad \mathcal{C}(0, 1) \simeq \mathbb{C}, \tag{8.20}$$

This is because the Clifford algebra is spanned by 1 and e, subject to $e^2 = \pm 1$. The second case is clear if we identify e with $\mathrm{i} = \sqrt{-1}$; in the first case, we introduce the basis $f = \frac{1}{2}(1 + e)$, $g = \frac{1}{2}(1 - e)$, which satisfies $f^2 = f$, $g^2 = g$, $fg = gf = 0$ to produce the decomposition. These are sometimes referred to as the *hyperbolic complex numbers*.

The three two-dimensional Clifford algebras are

$$\mathcal{C}(2, 0) \simeq \mathcal{C}(1, 1) \simeq M(2, \mathbb{R}), \qquad \mathcal{C}(0, 2) \simeq \mathbb{H}, \tag{8.21}$$

where $M(2, \mathbb{R})$ denotes the algebra of real 2×2 matrices, and $\mathbb{H}$ denotes the quaternions. This is because the Clifford algebra is spanned by $1, e_1, e_2, e_1 e_2$, subject to the relations $e_i^2 = \pm 1$. If the metric is negative definite, these can be identified with the quaternions $1, \mathrm{i}, \mathrm{j}, \mathrm{k}$. Alternatively, one can represent these generators by multiples of the Pauli matrices (7.32), namely $\sigma_0, \mathrm{i}\sigma_x, \mathrm{i}\sigma_y, -\mathrm{i}\sigma_z$. In the other two cases, we identify the basis elements with real 2×2 matrices, according to the following table:

	1	e_1	e_2	$e_1 e_2$
$\mathcal{C}(2, 0):$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$
$\mathcal{C}(1, 1):$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$
$\mathcal{C}(0, 2):$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -\mathrm{i} & 0 \\ 0 & \mathrm{i} \end{pmatrix}.$

For higher dimensions, the classification of real Clifford algebras is based on the following key result.

Theorem 8.1. *As algebras,*

$$\begin{aligned} \mathcal{C}(k + 2, l) &\simeq \mathcal{C}(2, 0) \otimes \mathcal{C}(l, k), \\ \mathcal{C}(k + 1, l + 1) &\simeq \mathcal{C}(1, 1) \otimes \mathcal{C}(k, l), \\ \mathcal{C}(k, l + 2) &\simeq \mathcal{C}(0, 2) \otimes \mathcal{C}(l, k). \end{aligned} \tag{8.22}$$

Proof: Let Q_1 be a quadratic form on $\mathbb{R}^2$, and let e_1, e_2 be an orthonormal basis. Let Q_2 be a quadratic form on $\mathbb{R}^{n-2}$. We identify $\mathbb{R}^n = \mathbb{R}^2 \oplus \mathbb{R}^{n-2}$, and write $z = x + y \in \mathbb{R}^n$, where $x = a e_1 + b e_2$ and $y \in \mathbb{R}^{n-2}$. Define the linear map

$$A \colon \mathbb{R}^n \longrightarrow \mathcal{C}(Q_1) \otimes \mathcal{C}(Q_2), \qquad A(x + y) = x \otimes 1 + (e_1 e_2) \otimes y. \qquad (8.23)$$

The object then is to extend A to be a Clifford algebra isomorphism from $\mathcal{C}(Q)$ to $\mathcal{C}(Q_1) \otimes \mathcal{C}(Q_2)$, by defining a suitable quadratic form Q on $\mathbb{R}^n$, so that the above decomposition is orthogonal with respect to Q. Thus we need to check that A maps the ideal $\mathcal{I}(Q)$ to 0. Now $A(z^2) = A(Q(z)) = Q(z)$. On the other hand,

$$A(z)^2 = x^2 \otimes 1 - (e_1^2 e_2^2) \otimes y^2 + (e_1 e_2 x + x e_1 e_2) \otimes y = x^2 - Q_1(e_1) Q_1(e_2) y^2,$$

where the third summand vanishes owing to the orthogonality of e_1, e_2 and the fact that x is a linear combination of them. Therefore, A will extend to an isomorphism of Clifford algebras provided we define

$$Q(z) = Q_1(x) - Q_1(e_1) Q_1(e_2) Q_2(y).$$

Thus if Q_1 is indefinite, then $Q = Q_1 + Q_2$, while if Q_1 is definite, then $Q = Q_1 - Q_2$. This suffices to prove the result. *Q.E.D.*

Example 8.2. For the case of $\mathbb{R}^3$ with positive definite inner product,

$$\mathcal{C}(3, 0) \simeq \mathcal{C}(2, 0) \otimes \mathcal{C}(0, 1) \simeq M(2, \mathbb{R}) \otimes \mathbb{C} \simeq M(2, \mathbb{C}).$$

Using the above representation of $\mathcal{C}(2, 0)$ and the form (8.23) of the map A, we find an explicit representation to be given by the Pauli matrices

$$1 \overset{\sim}{\longrightarrow} 1 \otimes 1 \overset{\sim}{\longrightarrow} \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad\qquad e_1 \overset{\sim}{\longrightarrow} e_1 \otimes 1 \overset{\sim}{\longrightarrow} \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$e_2 \overset{\sim}{\longrightarrow} e_2 \otimes 1 \overset{\sim}{\longrightarrow} \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad\qquad e_3 \overset{\sim}{\longrightarrow} e_1 e_2 \otimes 1 \overset{\sim}{\longrightarrow} \sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

For example, for the Clifford algebra on anti-Minkowski space, i.e., for the metric $(+ + + -)$, we have

$$\mathcal{C}(3, 1) \simeq \mathcal{C}(2, 0) \otimes \mathcal{C}(1, 1) \simeq M(2, \mathbb{R}) \otimes M(2, \mathbb{R}) \simeq M(4, \mathbb{R}),$$

is isomorphic to the algebra of 4×4 real matrices. An explicit form of this isomorphism is given by the proof of Theorem 8.1, i.e., if we let e_1, e_2 be a basis for $\mathbb{R}^2$ with the positive definite metric, and e_1', e_2' be a basis for $\mathbb{R}^2$ with the indefinite metric, then the above isomorphism is

$$e_1 \longrightarrow e_1 \otimes 1, \quad e_2 \longrightarrow e_2 \otimes 1, \quad e_3 \longrightarrow (e_1 e_2) \otimes e_2', \quad e_4 \longrightarrow (e_1 e_2) \otimes e_1'.$$

Using the earlier representations for $\mathcal{C}(2,0)$ and $\mathcal{C}(1,1)$ as $M(2,\mathbb{R})$, we find the associated representation of $\mathcal{C}(3,1)$ to be

$$e_1 \overset{\sim}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \qquad e_2 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$e_3 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \qquad e_4 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.$$

On the other hand for the Clifford algebra on Minkowski space, we have

$$\mathcal{C}(1,3) \simeq \mathcal{C}(0,2) \otimes \mathcal{C}(1,1) \simeq \mathbb{H} \otimes M(2,\mathbb{R}) \simeq M(2,\mathbb{H}),$$

is isomorphic to the algebra of 2×2 quaternionic matrices

$$e_1 \overset{\sim}{\longrightarrow} \mathrm{i} \otimes 1, \qquad e_2 \overset{\sim}{\longrightarrow} \mathrm{j} \otimes 1, \qquad e_3 \overset{\sim}{\longrightarrow} \mathrm{k} \otimes e_1', \qquad e_0 \overset{\sim}{\longrightarrow} \mathrm{k} \otimes e_2',$$

$$e_1 \overset{\sim}{\longrightarrow} \begin{pmatrix} \mathrm{i} & 0 \\ 0 & \mathrm{i} \end{pmatrix}, \qquad e_2 \overset{\sim}{\longrightarrow} \begin{pmatrix} \mathrm{j} & 0 \\ 0 & \mathrm{j} \end{pmatrix}, \qquad e_3 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \mathrm{k} \\ \mathrm{k} & 0 \end{pmatrix}, \qquad e_0 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \mathrm{k} \\ -\mathrm{k} & 0 \end{pmatrix}.$$

If we use the isomorphism of $\mathbb{H}$ with a subalgebra of $M(2,\mathbb{C})$ as given above, this leads to one form for the Dirac matrices

$$e_1 \overset{\sim}{\longrightarrow} \begin{pmatrix} \mathrm{i}\,\sigma_x & 0 \\ 0 & \mathrm{i}\,\sigma_x \end{pmatrix}, \qquad e_2 \overset{\sim}{\longrightarrow} \begin{pmatrix} \mathrm{i}\,\sigma_y & 0 \\ 0 & \mathrm{i}\,\sigma_y \end{pmatrix},$$

$$e_3 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & -\mathrm{i}\,\sigma_z \\ -\mathrm{i}\,\sigma_z & 0 \end{pmatrix}, \qquad e_0 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & -\mathrm{i}\,\sigma_z \\ \mathrm{i}\,\sigma_z & 0 \end{pmatrix}. \tag{8.24}$$

Alternative representations include the *Dirac form*

$$e_1 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix}, \qquad e_2 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix},$$

$$e_3 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}, \qquad e_0 \overset{\sim}{\longrightarrow} \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \tag{8.25}$$

and the *Weyl form*

$$e_1 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & -\sigma_x \\ \sigma_x & 0 \end{pmatrix}, \qquad e_2 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & -\sigma_y \\ \sigma_y & 0 \end{pmatrix},$$

$$e_3 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & -\sigma_z \\ \sigma_z & 0 \end{pmatrix}, \qquad e_0 \overset{\sim}{\longrightarrow} \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}. \tag{8.26}$$

For many purposes, it is the last form which is the most convenient, as the Dirac matrices exhibit a common block structure.

All these different representations are really equivalent. This is because there is only one irreducible representation of the algebra of $n \times n$ matrices.

Lemma 8.3. *Every irreducible representation of the matrix algebra $M(n)$ by $n \times n$ matrices is given by $\rho(A) = S A S^{-1}$.*

Idea of proof: Given a representation ρ on V, let $X : \mathbb{R}^n \longrightarrow V$ be any linear map. Set $S = \sum \rho(E_{ij}) X E_{ij}$, and show that S intertwines the two representations: $\rho(A) S = S A$, hence by Schur's Lemma 7.5, S must be a multiple of the identity. We then choose X so that $S \neq 0$. $\hspace{6cm}$ Q.E.D.

Another interesting example of a Clifford algebra is

$$\mathcal{C}(2,2) \simeq \mathcal{C}(2,0) \otimes \mathcal{C}(2,0) \simeq M(2,\mathbb{R}) \otimes M(2,\mathbb{R}) \simeq M(4,\mathbb{R}),$$
$$\simeq \mathcal{C}(0,2) \otimes \mathcal{C}(0,2) \simeq \mathbb{H} \otimes \mathbb{H},$$

leading to the isomorphism

$$\mathbb{H} \otimes \mathbb{H} \simeq M(4,\mathbb{R}).$$

Theorem 8.1 implies that there is an eight-fold periodicity among Clifford algebras:

$$\mathcal{C}(k+8,l) \simeq \mathcal{C}(k,l) \otimes \mathcal{C}(8,0), \qquad \mathcal{C}(k,l+8) \simeq \mathcal{C}(k,l) \otimes \mathcal{C}(0,8), \qquad (8.27)$$

which lies at the foundation of some deep results in algebraic topology and K-theory, e.g., the Bott Periodicity Theorem, [**11**]. From this, we are able to construct a complete table of real Clifford algebras:

$k - l \mod 8$	$\mathcal{C}(k,l)$ where $k + l = 2m$ or $2m+1$
0	$M(2^m, \mathbb{R})$
1	$M(2^m, \mathbb{R}) \oplus M(2^m, \mathbb{R})$
2	$M(2^m, \mathbb{R})$
3	$M(2^m, \mathbb{C})$
4	$M(2^{m-1}, \mathbb{H})$
5	$M(2^{m-1}, \mathbb{H}) \oplus M(2^{m-1}, \mathbb{H})$
6	$M(2^{m-1}, \mathbb{H})$
7	$M(2^m, \mathbb{C})$

Note that these algebras are simple except for $k - l \equiv 1$ or $5 \mod 8$.

For complex Clifford algebras, there is just one non-degenerate $\mathcal{C}(n)$ for each dimension n, and the above results simplify:

$$\mathcal{C}(2m) \simeq M(2^m, \mathbb{C}), \qquad \mathcal{C}(2m+1) \simeq M(2^m, \mathbb{C}) \oplus M(2^m, \mathbb{C}), \qquad (8.28)$$

and so there is a periodicity of 2. Note the curious fact that the complex odd-dimensional Clifford algebras do not occur as real Clifford algebras!

The Clifford algebra $\mathcal{C}(Q)$ comes equipped with a natural grading into odd and even order parts, stemming from the grading of the tensor algebra, since the basic relations take even order tensors to even order tensors. We let $\mathcal{C}^+(Q)$ denote the even subalgebra, which includes the scalars, but not V itself.

Proposition 8.4. *As algebras,*

$$\mathcal{C}^+(p+1,q) \simeq \mathcal{C}(q,p), \qquad \mathcal{C}^+(p,q+1) \simeq \mathcal{C}(p,q), \tag{8.29}$$

Proof: Let Q_1 be a quadratic form on $\mathbb{R}^{n-1}$, and let Q_2 be a quadratic form on $\mathbb{R}$. Form the decomposition $\mathbb{R}^n = \mathbb{R}^{n-1} \oplus \mathbb{R}$, with e_n the normalized basis vector for $\mathbb{R}$. Define the linear map

$$A(x) = x \, e_n, \qquad A: \mathbb{R}^{n-1} \longrightarrow \mathcal{C}(Q),$$

where Q is a suitable quadratic form on $\mathbb{R}^n$, making the decomposition orthogonal. Then

$$A(x^2) = A(Q_1(x)) = Q_1(x).$$

On the other hand,
$$A(x)^2 = -x^2 \, e_n^2 = -Q(e_n) Q(x).$$

Therefore, A will extend to a map between Clifford algebras provided we define

$$Q(x+y) = Q_2(y) - Q_2(e_n) Q_1(x), \qquad x \in \mathbb{R}^{n-1}, \qquad y = \lambda e_n.$$

In other words, if $Q(e_n) = +1$, then Q_1 has the opposite sign to Q, while if $Q(e_n) = -1$, then Q_2 has the same sign as Q. $\hspace{2cm}$ Q.E.D.

For example, we find

$$\mathcal{C}^+(3,1) \simeq \mathcal{C}^+(1,3) \simeq \mathcal{C}(3,0) \simeq M(2,\mathbb{C}),$$

even though $\mathcal{C}(3,1)$ and $\mathcal{C}(1,3)$ are not isomorphic. In general, for $k > 1$,

$$\mathcal{C}(k,l) \simeq \mathcal{C}^+(k,l+1) \simeq \mathcal{C}(l+1,k-1),$$

which can be checked in the table. For complex Clifford algebras, the same proof shows that $\mathcal{C}^+(n) \simeq \mathcal{C}(n-1)$.

Clifford Groups and Spinors

There are a couple of important maps on any Clifford algebra. The *transpose* is defined by

$$\lambda^T = \lambda, \qquad \lambda \in K, \qquad v^T = v, \qquad v \in V, \qquad (xy)^T = y^T x^T, \qquad x, y \in \mathcal{C}(Q).$$

The automorphism

$$\alpha(\lambda) = \lambda, \qquad \lambda \in K, \qquad \alpha(v) = -v, \qquad v \in V, \qquad \alpha(xy) = \alpha(x) \, \alpha(y), \qquad x, y \in \mathcal{C}(Q),$$

is also important. Note that

$$\mathcal{C}^+(Q) = \{\, x \mid \alpha(x) = x \,\}, \qquad \mathcal{C}^-(Q) = \{\, x \mid \alpha(x) = -x \,\}.$$

The *Clifford group* $\mathcal{G}(Q)$ associated with a Clifford algebra $\mathcal{C}(Q)$ consists of all invertible elements $g \in \mathcal{C}(Q)$ such that

$$\rho(g) \, v \equiv \alpha(g) \, v \, g^{-1} \in V \qquad \text{for all} \qquad v \in V.$$

Note that ρ defines a representation of G on the vector space V.

We let $\mathrm{O}(Q)$ denote the group of *linear isometries* of the (nondegenerate) quadratic form Q, meaning those $A \in \mathrm{GL}(V)$ that preserve it: $Q(Av) = Q(v)$ for all $v \in V$, and $\mathrm{SO}(Q)$ those of unit determinant. If Q is real, of signature (k, l), then $\mathrm{O}(Q) \simeq \mathrm{O}(k, l)$ and similarly for $\mathrm{SO}(Q)$.

Lemma 8.5. *Given $g \in \mathcal{G}(Q)$, the map $\rho(g)$ is a linear isometry of the underlying quadratic form Q.*

Proof: Note that since $\rho(g)v \in V$,

$$\alpha(\rho(g)v) = -\rho(g)v,$$

and hence

$$\rho(g)v = \alpha(g)vg^{-1} = -\alpha(\alpha(g)vg^{-1}) = -g\alpha(v)\alpha(g^{-1}) = gv\alpha(g)^{-1}.$$

Therefore, since $Q(v)$ is a scalar,

$$Q(\rho(g)v) = (\rho(g)v)^2 = (\alpha(g)vg^{-1})(gv\alpha(g^{-1})) = \alpha(g)v^2\alpha(g)^{-1}$$
$$= \alpha(g)Q(v)\alpha(g)^{-1} = Q(v),$$

which shows that $\rho(g)$ is a linear isometry of Q.   Q.E.D.

We also define the restricted Clifford group, denote $\mathcal{G}^+(Q)$, to be the intersection of $\mathcal{G}(Q)$ with $\mathcal{C}^+(Q)$, the space of even rank elements of the Clifford algebra.

Theorem 8.6. *We have*

$$\rho(\mathcal{G}(Q)) = \mathrm{O}(Q), \qquad \rho(\mathcal{G}^+(Q)) = \mathrm{SO}(Q), \qquad \ker \rho = \rho^{-1}(\mathrm{I}) = K^* \equiv K \setminus 0. \quad (8.30)$$

Proof: Note first that $V \cap \mathcal{G}(Q)$ consists of all non-isotropic vectors $w \in V$, since the inverse of w is $w/Q(w)$ if $Q(w) \neq 0$, whereas $w^2 = 0$ if $Q(w) = 0$, so $w \in G$. Now, for $w \in V \cap \mathcal{G}(Q)$,

$$\rho(w)v = -w^{-1}vw = -\frac{wvw}{Q(w)} = -\frac{w(2Q(w, v) - wv)}{Q(w)} = w - \frac{2Q(w, v)}{Q(w)}v.$$

The last map is just the reflection in the hyperplane Q-orthogonal to w. The first two results now follow from a Theorem of Cartan and Dieudonné that states that every orthogonal transformation of an n-dimensional space can be written as the product of $m \leq n$ such reflections.

Next, to show $\ker \rho = K^*$, we need to characterize those $g \in G$ such that

$$\alpha(g)v = vg, \qquad \text{for all} \qquad v \in V.$$

Write $g = g_+ + g_-$, where $g_\pm \in \mathcal{C}^\pm(Q)$. Then $\alpha(g) = g_+ - g_-$. Thus the above condition becomes

$$g_+v = vg_+, \qquad g_-v = -vg_-, \qquad \text{for all} \qquad v \in V.$$

Introduce an orthonormal basis e_j of V relative to Q. Given any e_j, write

$$g_+ = a_+ + e_j b_-, \qquad \text{where} \qquad a_+ \in \mathcal{C}^+(Q), \qquad b_- \in \mathcal{C}^-(Q),$$

are such that they do not involve the basis vector e_j. Then

$$g_+ e_j = (a_+ + e_j b_-) \, e_j = a_+ e_j - Q(e_j) b_-,$$

whereas

$$e_j g_+ = a_+ e_j + Q(e_j) \, b_-.$$

Therefore, the above condition proves that $b = 0$, so the basis vector e_j does not appear in g_+. The only elements with no basis vectors in them at all are elements of K, so $g_+ \in K^*$. A similar argument proves that g_- has no basis vectors in it, so $g_- = 0$. $\hspace{2cm}$ Q.E.D.

Given $x \in \mathcal{C}(Q)$, define

$$N(x) = \alpha(x)^T x. \tag{8.31}$$

For example, if $x = v_1 v_2 \cdots v_k$ is decomposable, then

$$N(v_1 v_2 \ \cdots \ v_k) = (-1)^k (v_k \ \cdots \ v_2 v_1) \, (v_1 v_2 \ \cdots \ v_k) = (-1)^k Q(v_1) Q(v_2) \ \cdots \ Q(v_k)$$

is an element of K. However, not every element has $N(x) \in K$; for example

$$N(e_1 e_2 + e_3 e_4) = (e_2 e_1 + e_4 e_3) \, (e_1 e_2 + e_3 e_4) = Q(e_1) \, Q(e_2) + Q(e_3) \, Q(e_4) - 2 e_1 e_2 e_3 e_4.$$

Lemma 8.7. *The function (8.31) defines a map $N : \mathcal{G}(Q) \longrightarrow K^*$ called the induced spin norm on the Clifford group G.*

Proof: Using the previous result, we need only show that $N(g) \in \ker \rho$. Note that

$$\alpha(N(g)) = g^T \alpha(g).$$

Now

$$\alpha(g) \, v \, g^{-1} = \rho(g) \, v = (\rho(g) \, v)^T = g^{-T} v \, \alpha(g)^T,$$

hence

$$v = g^T \alpha(g) \, v (\alpha(g)^T g)^{-1} = \alpha(N(g)) \, v N(g)^{-1} = \rho(N(g)) \, v,$$

for all $v \in V$. Therefore $\rho(N(g)) = \mathrm{I}$. $\hspace{3cm}$ Q.E.D.

Note that if g is any element with $N(g) \in K^*$, then we have the useful formula

$$g^{-1} = \frac{\alpha(g)^T}{N(g)}.$$

If $g \in G$, then this gives

$$\rho(g) \, v = g \, v \, \alpha(g^{-1}) = \frac{g \, v \, g^T}{N(g)}.$$

It is not usually true, however, that $N(g) \in K^*$ implies that $g \in \mathcal{G}(Q)$ or $\mathcal{G}^+(Q)$. For instance, in $\mathcal{C}(2,0)$,

$$N(e_1 + \lambda \, e_1 e_2) = (-e_1 + \lambda \, e_2 e_1) \, (e_1 + \lambda \, e_1 e_2) = \lambda^2 - 1,$$

which is nonzero for $\lambda \neq \pm 1$, but

$$\rho(e_1 + \lambda e_1 e_2) e_1 = (-e_1 + \lambda e_2 e_1) e_1 (e_1 + \lambda e_1 e_2) = (\lambda^2 - 1) e_1 - 2\lambda e_1 e_2.$$

For a real vector space, define the groups

$$\begin{aligned}
\mathrm{Pin}(Q) &= \{\, g \in \mathcal{G}(Q) \mid N(g) = \pm 1 \,\}, \\
\mathrm{Spin}(Q) &= \{\, g \in \mathcal{G}^+(Q) \mid N(g) = \pm 1 \,\}.
\end{aligned} \tag{8.32}$$

Theorem 8.8. *The representations*

$$\rho : \mathrm{Pin}(Q) \longrightarrow \mathrm{O}(Q), \qquad \rho^S : \mathrm{Spin}(Q) \longrightarrow \mathrm{SO}(Q), \tag{8.33}$$

are two-to-one coverings, with $\rho(A) = \rho(B)$ if and only if $A = \pm B$. Moreover, ρ^S induces a nontrivial double covering of $\mathrm{SO}(Q)$ as long as $\dim V \geq 3$, or $\dim V = 2$ and Q is definite.

Proof: In the first case, it suffices to note that $\mathrm{Pin}(Q)$ includes all vectors $v \in V$ with $Q(v) = \pm 1$, and $\rho(v)$ is just the reflection through the hyperplane orthogonal to V. Moreover, $\rho(A) = \mathrm{I}$ if and only if $A = \lambda \in \mathbb{R}$, and $N(A) = \lambda = \pm 1$. For the second result we proceed the same way, using the fact that every element of $\mathrm{SO}(Q)$ can be written as a product of an even number of reflections, since each reflection has determinant -1.

Note that these imply that every element of $\mathrm{Pin}(Q)$ can be written as $v_1 v_2 \cdots v_k$ for $k \leq n$, where $Q(v_i) = \pm 1$, i.e., $\mathrm{Pin}(Q)$ is generated by the unit "sphere" in V. Similarly for $\mathrm{Spin}(Q)$, except that k must be even.

The nontriviality of the covering follows from the fact that we can connect 1 and -1 by the curve

$$\cos t + e_1 e_2 \sin t, \qquad 0 \leq t \leq \pi,$$

which is contained in $\mathrm{Spin}(Q)$ as long as $Q(e_1) Q(e_2) > 0$, $Q(e_1, e_2) = 0$. Therefore, $\mathrm{Spin}(Q)$ is the simply connected covering group for $\mathrm{SO}(Q)$.   *Q.E.D.*

Example 8.9. Consider the case $V = \mathbb{R}^3$, with positive definite inner product. Then $\mathcal{C}^+(3,0) \simeq \mathbb{H}$, and consists of all elements of the form

$$x = a + b e_1 e_2 + c e_3 e_1 + d e_2 e_3.$$

Note first that $\rho(x) v \in V$ for all $v \in V$. We find

$$N(x) = (a + b e_2 e_1 + c e_1 e_3 + d e_3 e_2)(a + b e_1 e_2 + c e_3 e_1 + d e_2 e_3) = a^2 + b^2 + c^2 + d^2,$$

hence every nonzero element is in $\mathcal{G}^+(3,0)$. The representation $\rho : \mathbb{H}^* \longrightarrow \mathrm{SO}(3)$ reproduces the well known correspondence between quaternions and rotations. Alternatively, we can identify $1, e_2 e_3, e_3 e_1, e_1 e_2$ respectively with the unitary matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathrm{i}\,\sigma_x = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \mathrm{i}\,\sigma_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \mathrm{i}\,\sigma_z = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}. \tag{8.34}$$

Also note that $\mathcal{C}^+(3,0)$ is the same as the real matrix algebra spanned by the Pauli matrices, and the $\mathcal{G}^+(3,0)$ is the subgroup of invertible such matrices since

$$\det(a\sigma_0 + b\sigma_x + c\sigma_y + d\sigma_z) = a^2 + b^2 + c^2 + d^2.$$

Furthermore, we can identify the double covering group $\mathrm{Spin}(3)$ with the group $\mathrm{SU}(2)$ as before.

In the case of Minkowski space, the general element of $\mathcal{C}^+(1,3)$ has the form

$$x = a + b\,e_1 e_2 + c\,e_1 e_3 + d\,e_2 e_3 + e\,e_0 e_1 + f\,e_0 e_2 + g\,e_0 e_3 + h\,e_0 e_1 e_2 e_3.$$

Then

$$N(x) = (a^2 + b^2 + c^2 + d^2 - e^2 - f^2 - g^2 - h^2) + 2(a\,h - b\,g + c\,f - d\,e)e_0 e_1 e_2 e_3.$$

We also find that $\rho(x)\,v \in V$ for all $v \in V$ if and only if $N(x) \in \mathbb{R}$. Therefore, $\mathcal{G}^+(1,3)$ consists of those elements with

$$a^2 + b^2 + c^2 + d^2 \neq e^2 + f^2 + g^2 + h^2, \qquad a\,h - b\,g + c\,f - d\,e = 0.$$

We can identify $\mathcal{C}^+(1,3)$ with the algebra of 2×2 complex matrices. The explicit isomorphism is given by identifying $1, e_2 e_3, e_1 e_3, e_1 e_2$ with the matrices (8.34) as before, along with identifying $e_0 e_1, e_0 e_2, e_0 e_3, e_0 e_1 e_2 e_3$, respectively, with the matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad -\sigma_y = \begin{pmatrix} 0 & \mathrm{i} \\ -\mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathrm{i}\,\sigma_0 = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & \mathrm{i} \end{pmatrix}.$$

Under this identification

$$N(x) = \mathrm{Re}\,(\det X) + \mathrm{Im}\,(\det X)\,e_0 e_1 e_2 e_3,$$

we see that $\mathcal{G}^+(1,3)$ is the set of such matrices which have real, non-zero determinant. Furthermore, we can identify $\mathrm{Spin}(1,3)$ with the group

$$\{\,X \mid \det X = \pm 1\,\} = \mathrm{SL}(2,\mathbb{C}) \cup \mathrm{SL}^-(2,\mathbb{C}),$$

which is just two disconnected copies of $\mathrm{SL}(2,\mathbb{C})$. The representation ρ takes $\mathcal{G}^+(1,3)$ to $\mathrm{SO}(1,3)$, which also has two connected components. Finally, $\mathcal{G}(1,3)$ maps onto the full Lorentz group; the formulas are left to the reader.

Representations of the Lorentz Group

References: [**29**], [**37**], [**38**].

The (finite-dimensional) real representation theory of $\mathrm{SL}(2,\mathbb{C})$ proceeds as follows. First note that if we consider $\mathrm{SL}(2,\mathbb{C})$ as a complex Lie group, then its complex analytic representations will correspond to the real analytic representations of any real form. In particular, we can use the tensor representation theory of $\mathrm{SL}(2,\mathbb{R})$, or, better the representation theory of the compact Lie group $\mathrm{SU}(2)$, both of which are real forms of $\mathrm{SL}(2,\mathbb{C})$. Thus the complex finite-dimensional representations of $\mathrm{SL}(2,\mathbb{C})$ are indexed by a half-integer $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$. We can explicitly realize these representations on the space of homogeneous polynomials $p(\mathbf{z})$, where $\mathbf{z} = (z_1, z_2) \in \mathbb{C}^2$, of degree $2j$ by

$$\rho^j(A)\,p(\mathbf{z}) = p(A^{-1}\mathbf{z}). \tag{8.35}$$

It is a preferable to use homogeneous (projective) coordinates, so write

$$f(z) = p(z,1) = z_2^{-2j}\,p(z_1, z_2), \qquad \text{where} \qquad z = z_1/z_2,$$

so that f is an inhomogeneous polynomial in z having degree $2j$. The representation ρ^j becomes the linear fractional representation

$$\rho^j(A)f(z) = (\gamma z + \delta)^{2j} f\left(\frac{\alpha z + \beta}{\gamma z + \delta}\right), \qquad A^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{8.36}$$

Next consider the complex representations of $\mathrm{SL}(2,\mathbb{C})$ considered as a real Lie group. In general if $\mathfrak{g}$ is any complex m-dimensional Lie algebra with real form $\mathfrak{g}_\mathbb{R}$, let $\widehat{\mathfrak{g}}$ denote it considered as just a real Lie algebra, of dimension $2m$. If $A_1, \ldots, A_m$ form a basis for $\mathfrak{g}_\mathbb{R}$, then they also form a complex basis for $\mathfrak{g}$. Therefore, $A_1, \ldots, A_m, B_1 = \mathrm{i}\, A_1, \ldots, B_m = \mathrm{i}\, A_m$, form a real basis for $\widehat{\mathfrak{g}} \simeq \mathfrak{g}_\mathbb{R} \oplus \mathrm{i}\, \mathfrak{g}_\mathbb{R}$. The complexification of $\widehat{\mathfrak{g}}$, denoted as $\widehat{\mathfrak{g}}_\mathbb{C}$ is therefore isomorphic to $\mathfrak{g} \oplus \mathfrak{g}$, and we can use $C_k = A_k + \mathrm{i}\, B_k$ and $C_k^* = A_k - \mathrm{i}\, B_k$ as a basis implementing this decomposition. Therefore, the analytic irreducible representations of $\widehat{\mathfrak{g}}$ are realized as tensor products of irreducible representations of $\mathfrak{g}$.

In our case, this implies that the Lorentz algebra $\mathfrak{so}(1,3)$ is a real form of the complex Lie algebra $\mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$, whose other real forms include $\mathfrak{su}(2) \oplus \mathfrak{su}(2)$. The irreducible representations of $\mathfrak{so}(1,3)$ are parametrized by two integers or half integers:

$$\rho^{j,k} = \rho^j \otimes \rho^k, \qquad \text{where} \qquad j, k = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \ \ldots \ .$$

It is usually more convenient to identify the second factor with its complex conjugate (since the generators are complex conjugates of the generators for the first factor), and so we write

$$\rho^{j,k} = \rho^j \otimes \overline{\rho}^k, \qquad \text{where} \qquad j, k = 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots .$$

Note that the dimension of $\rho^{j,k}$ is $(2j+1)(2k+1)$. The complex conjugate of the representation $\rho^{j,k}$ is the representation $\rho^{k,j}$. In particular, the only real representations are the ones where $j = k$.

The Clebsch-Gordan series (7.63) for tensor products of these representations are readily computed from those of $\mathfrak{su}(2)$, so

$$\rho^{j,k} \otimes \rho^{j',k'} = \bigoplus_{i=|j-j'|}^{j+j'} \ \bigoplus_{l=|k-k'|}^{k+k'} \rho^{i,l}. \tag{8.37}$$

The restriction of $\rho^{j,k}$ to the subgroup $\mathrm{SU}(2) \subset \mathrm{SL}(2,\mathbb{C})$ decomposes as $\rho^j \otimes \rho^k$, since $\mathrm{SU}(2)$ is essentially the diagonal subgroup for the direct sum decomposition.

The standard representation of the Lorenz group $L = \mathrm{SO}(1,3)$ on four-dimensional Minkowski space is equivalent to the irreducible representation $\rho^{1/2,1/2}$. One way to verify this (see below for an explicit approach) is to note that this representation is clearly irreducible. The only four-dimensional representations of $\mathfrak{so}(1,3)$ are $\rho^{1/2,1/2}, \rho^{3/2,0}$, and $\rho^{0,3/2}$. But on Minkowski space, the subgroup $\mathrm{SO}(3)$ (which corresponds to $\mathrm{SU}(2)$ in the double covering) decomposes into a physical representation and a trivial representation on the time axis, i.e., has the representation $\rho^1 \oplus \rho^0$. The only one of the above three representations which has the same property is $\rho^{1/2,1/2}$; the other two remain irreducible upon restriction to $\mathrm{SU}(2)$.

In view of (8.35), the real representation $\rho^{j,k}$ of $\mathrm{SL}(2,\mathbb{C})$ can be realized on the space of homogeneous polynomials $p(\mathbf{z},\overline{\mathbf{z}})$ of degree $2j$ in $\mathbf{z} = (z_1, z_2)$ and $2k$ in $\overline{\mathbf{z}} = (\overline{z}_1, \overline{z}_2)$ by

$$\rho^{j,k}(A)\,p(\mathbf{z},\overline{\mathbf{z}}) = p(A^{-1}\mathbf{z}, \overline{A}^{-1}\overline{\mathbf{z}}). \tag{8.38}$$

As above, it is preferable to use homogeneous coordinates, and write

$$f(z,\overline{z}) = p(z, 1, \overline{z}, 1) = z_2^{-2j}\overline{z}_2^{-2k}p(z_1, z_2, \overline{z}_1, \overline{z}_2), \qquad \text{where} \qquad z = z_1/z_2,$$

so that f is an inhomogeneous polynomial in $z, \overline{z}$, having degree $(2j, 2k)$. The representation $\rho^{j,k}$ becomes the linear fractional representation

$$\rho^{j,k}(A)\,f(z,\overline{z}) = (\gamma z + \delta)^{2j}\,(\overline{\gamma z + \delta})^{2k} f\left(\frac{\alpha z + \beta}{\gamma z + \delta}, \frac{\overline{\alpha z + \beta}}{\overline{\gamma z + \delta}} \right), \qquad A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{8.39}$$

Thus, to check the earlier assertion about the representation $\rho^{1/2,1/2}$ being that of the usual Lorentz group, we use the space of complex bilinear polynomials

$$a\,z_1\overline{z}_1 + b\,z_1\overline{z}_2 + c\,z_2\overline{z}_1 + d\,z_2\overline{z}_2.$$

as the basic representation space $\rho^{1/2,0} \otimes \rho^{0,1/2}$. This can be identified with the space of 2×2 complex matrices X, which transform according to $X \longmapsto \overline{A}^T X A$ under $\mathrm{SL}(2,\mathbb{C})$. Therefore the determinant is invariant. The subspace $\mathcal{H}$ of Hermitian matrices is the real form of this representation, which is equivalent to its complex conjugate, and hence real; we use the earlier formulation (8.8) to identify this quantity with the Minkowski metric, whereby

$$(a, b, c, d) \qquad \text{corresponds to} \qquad (c\,t + z, x + \mathrm{i}\,y, x - \mathrm{i}\,y, c\,t - z).$$

Since $\mathrm{SL}(2,\mathbb{C})$ is not compact, not all of its irreducible representations are finite-dimensional. For example, the analogous representation

$$\rho^{j,k}(A)\,f(z,\overline{z}) = (\gamma z + \delta)^{j + \mathrm{i}\,c - 1}(\overline{\gamma z + \delta})^{-j + \mathrm{i}\,c - 1} f\left(\frac{\alpha z + \beta}{\gamma z + \delta}, \frac{\overline{\alpha z + \beta}}{\overline{\gamma z + \delta}} \right), \tag{8.40}$$

where $j = 0, \frac{1}{2}, 1, \ldots$, and c is any real number, is defined on the Hilbert space of L^2 functions of the real variables x, y, where $z = x + \mathrm{i}\,y$, and defines an irreducible infinite-dimensional unitary representation. The only finite-dimensional unitary representation is the trivial representation $\rho^{0,0}$. A complete classification of these and other representations can be found in [**38**]. In general, the representations of the group $\mathrm{SL}(2,\mathbb{C})$ are double-valued representations of the proper Lorentz group $L_+^\uparrow$. Note that this will be the case if and only if $\rho(-\mathrm{I}) = \mathrm{I}$, where $-\mathrm{I}$ is the element of $\mathrm{SL}(2,\mathbb{C})$. From the above realization, this clearly holds if and only if $j + k$ is an integer.

Next consider the group $L^\uparrow$ generated by $L_+^\uparrow$ and the parity reversal matrix $P = \mathrm{diag}\,(1, -1, -1, -1)$. Recall that P corresponds to the transformation $S \circ C$ acting on the space $\mathcal{H}$ of Hermitian matrices. On the associated representation spaces, this transformation is equivalent to the operation

$$z_1 \longmapsto \overline{z}_2, \qquad z_2 \longmapsto -\overline{z}_1,$$

which maps the representation $\rho^{j,k}$ to the conjugate representation $\rho^{k,j}$. If $j \neq k$, then we form an irreducible representation by summing these two representations: $\rho^{j,k} \oplus \rho^{k,j}$; it can be shown that, for $j \neq k$, this is an irreducible representation of $L^\uparrow$. If $j = k$, then this representation decomposes into two irreducible representations; the representation space splits into invariant subspaces

$$V^{j,j} \oplus V^{j,j} = V_+^{j,j} \oplus V_-^{j,j}, \qquad V_\pm^{j,j} = \left\{ (v, \pm v) \mid v \in V^{j,j} \right\}.$$

The two resulting representations, $\rho_+^{j,j}$ and $\rho_-^{j,j}$ are then irreducible representations of the simply connected covering group for $L^\uparrow$. (Unfortunately, our use of $\pm$ does not quite agree with the standard use; see [**37**].) These representations are single-valued if and only if $j + k$ is an integer. The usual representation of $L^\uparrow$ on Minkowski space is equivalent to $\rho_+^{1/2,1/2}$, since we identify a 2×2 matrix with the bilinear polynomial

$$a\,(z_1 \overline{z}_1 + w_1 \overline{w}_1) + b\,(z_1 \overline{z}_2 + w_1 \overline{w}_2) + c\,(z_2 \overline{z}_1 + w_2 \overline{w}_1) + d\,(z_2 \overline{z}_2 + w_2 \overline{w}_2),$$

where (a, b, c, d) represents $(ct + z, x + \mathrm{i}\,y, x - \mathrm{i}\,y, ct - z)$. The action of P, which then takes z_1 to $\overline{w}_2$ and z_2 to $-\overline{w}_1$, and vice versa, has the effect

$$(a, b, c, d) \longmapsto (d, -b, -c, a),$$

which is the same as the parity reversing map P on (t, x, y, z).

The full Lorentz group is obtained from the group $L^\uparrow$ by adding in total inversion $-\,\mathrm{I}$. Since this commutes with all elements, and its square is the identity, to each representation of $L^\uparrow$ there corresponds two representations of L, one where total inversion is represented by the identity I; the other where it is represented by $-\,\mathrm{I}$.

Relativistic Spinors

For the Lorentz group, one introduces two kinds of basic spinors: those corresponding to the standard representation $\rho^{1/2,0} = \rho^{1/2}$ of $\mathrm{SL}(2, \mathbb{C})$ on the two-dimensional vector space $\mathbb{C}^2$, and those corresponding to the complex conjugate representation $\rho^{0,1/2}$. These are then multiplied together in the obvious way to give spinors of higher rank. Traditionally, one uses regular spinor indices to denote the former and dotted spinor indices to denote the latter. Thus, the representation space of $\rho^{j,k}$ is the same as the space of completely symmetric spinors in j undotted and k dotted indices. In particular, ordinary vectors are written as spinors with one undotted and one dotted index. Therefore, two-dimensional spinors in the representation $\rho^{1/2,0}$ will have components (a_1, a_2) corresponding to the linear polynomial $a_1 z_1 + a_2 z_2$, whereas those in the representation $\rho^{0,1/2}$ will have the dotted components $(a_{\dot{1}}, a_{\dot{2}})$ corresponding to the linear polynomial $a_{\dot{1}} \overline{z}_1 + a_{\dot{2}} \overline{z}_2$. The ordinary vectors in Minkowski space will have the four components $(a_{1\dot{1}}, a_{1\dot{2}}, a_{2\dot{1}}, a_{2\dot{2}})$ corresponding to the bilinear polynomial

$$a_{1\dot{1}}\, z_1 \overline{z}_1 + a_{1\dot{2}}\, z_1 \overline{z}_2 + a_{2\dot{1}}\, z_2 \overline{z}_1 + a_{2\dot{2}}\, z_2 \overline{z}_2.$$

In terms of the ordinary x, y, z, t coordinates, we use the identification of this polynomial with a 2×2 complex matrix, and restrict to the Hermitian matrices to deduce

$$a_{1\dot{1}} = ct - z, \qquad a_{1\dot{2}} = x + \mathrm{i}\,y, \qquad a_{2\dot{1}} = x - \mathrm{i}\,y, \qquad a_{2\dot{2}} = ct + z. \qquad (8.41)$$

Lorentz–Invariant Equations

Since the fundamental symmetry group of special relativity is the Lorentz group (or its double covering), it is of great interest to classify equations which are invariant under it. In the present approach, we will look for first order equations of the form

$$(A^\mu P_\mu + B)\psi = 0, \qquad P_\mu = i\hbar\,\partial_\mu, \tag{8.42}$$

where A^μ, B are matrices, which are invariant under some representation of $\mathrm{SL}(2,\mathbb{C})$.

Note first that the differential operators P_μ transform under $\mathrm{SL}(2,\mathbb{C})$ according to the representation $\rho^{1/2,1/2}$. Indeed, if $\widetilde{\mathbf{q}} = A\mathbf{q}$ is any change of variables, then the derivatives transform according to the contragredient representation $\widetilde{\partial} = A^{-T}\partial$. Observe that if $A \in \mathrm{SL}(2,\mathbb{C})$, then

$$A^{-T} = JAJ^{-1} = JAJ, \qquad \text{where} \qquad J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Therefore, passing to the contragredient representation of $\mathrm{SL}(2,\mathbb{C})$ amounts to changing bases by

$$z_1 \longmapsto z_2, \qquad z_2 \longmapsto -z_1,$$

and similarly for the complex conjugate representation. Thus, in the spinor form of the representation $\rho^{j,k}$, the contragredient representation is obtained by interchanging the 1's and 2's, with appropriate plus or minus sign. In particular, the spinor form for the momenta are the operators

$$\partial_{1\dot{1}} = c\,\partial_t + \partial_z, \quad \partial_{1\dot{2}} = -\partial_x + i\,\partial_y, \quad \partial_{2\dot{1}} = -\partial_x - i\,\partial_y, \quad \partial_{2\dot{2}} = c\,\partial_t - \partial_z. \tag{8.43}$$

If the wave function ψ transforms according to the irreducible representation $\rho^{j,k}$, then the associated momenta $P_\mu\psi$ will transform according to the representation

$$\rho^{1/2,1/2}\otimes\rho^{j,k} = \rho^{j+1/2,k+1/2}\oplus\rho^{j-1/2,k+1/2}\oplus\rho^{j+1/2,k-1/2}\oplus\rho^{j-1/2,k-1/2}, \tag{8.44}$$

with obvious modifications if j or k is 0. Now if B is invertible, we can rewrite (8.42) in the form

$$K^\mu P_\mu\,\psi = \lambda\,\psi. \tag{8.45}$$

The parameter λ represents the mass. If we decompose ψ into irreducible representation components, we see that, for massive particles $\lambda > 0$, the equation (8.45) will couple the $\rho^{j,k}$ component of ψ to the components in the representations appearing in (8.44), and thus ψ cannot transform according to a single irreducible representation of $\mathfrak{sl}(2,\mathbb{C})$. We deduce that a Lorentz-invariant equation for a particle with mass will involve a combination of various representations. On the other hand, massless particles, i.e., when $\lambda = 0$, can be formed from a single irreducible representation.

For example, if we start out with ψ a scalar, i.e., transforming by the trivial representation $\rho^{0,0}$, then $P_\mu\psi = \psi_{,\mu}$ will transform according to the representation $\rho^{1/2,1/2}$. The components of $P_\nu\psi_{,\mu}$ will transform like $\rho^{1/2,1/2}\otimes\rho^{1/2,1/2}$, and to obtain a simple closed

system we need to single out the $\rho^{0,0}$ summand of this tensor product. This is done using the Clebsch-Gordan coefficients computed in (8.37). The result is the Klein-Gordon equation (8.13) written in the first order form

$$P_\mu \psi = \psi_{,\mu}, \qquad P^\mu \psi_{,\mu} = \lambda \psi. \tag{8.46}$$

This describes the simplest system for a particle of spin 0.

For a particle of spin $\frac{1}{2}$, we need to look at the representation $\rho^{1/2,0}$. To construct a Lorentz-invariant equation for a field ψ transforming according to this representation, we need to consider

$$\rho^{1/2,1/2} \otimes \rho^{1/2,0} = \rho^{1,1/2} \oplus \rho^{0,1/2}.$$

Thus, if we single out the $\rho^{0,1/2}$ component, we can obtain a massless two-component equation whose left hand side transforms according to $\rho^{0,1/2}$. This is the *Weyl equation*,

$$\sigma^\mu P_\mu \psi = 0, \tag{8.47}$$

where the σ^μ's are the Pauli matrices (7.32), which was originally rejected as a physical equation, since it does not conserve parity. However, it was subsequently vindicated when the neutrino was observed not to conserve parity, and is used there.

An electron, however, has spin $\frac{1}{2}$, but does conserve parity. Therefore, it should transform under the four-dimensional irreducible representation $\rho^{1/2,0} \oplus \rho^{0,1/2}$. Similar arguments lead to a system of two coupled sets of Weyl equations:

$$\sigma^\mu P_\mu \psi = \lambda \widetilde{\psi}, \qquad g^{\mu\nu} \sigma^\nu P_\mu \widetilde{\psi} = \lambda \psi, \tag{8.48}$$

where the λ's are equal to maintain invariance under space inversion. This is another form of the Dirac equation, based on the Weyl representation (8.26) for the Dirac matrices, i.e., this is equivalent to

$$\gamma^\mu P_\mu \boldsymbol{\psi} = m \boldsymbol{\psi},$$

where $\boldsymbol{\psi}$ represents the pair of spinors $(\psi, \widetilde{\psi})$. The individual components of $\boldsymbol{\psi}$ satisfy the Klein-Gordon equation (8.13).

Another important example of an invariant equation is Maxwell's equations for an electromagnetic field in a vacuum. The details can be found in the references.

Symmetries of the Dirac Equation

The Dirac equation (8.14) is invariant under the full Lorentz group. Indeed, by our results on the Clifford group, we know that the Lorentz group acts on the Dirac matrices, which generate the Clifford algebra, according to the standard representation, including the discrete symmetries of parity reversal, P, and time reversal, T. In addition, it also has the symmetry of charge conjugation, denoted by C.

Note that each Dirac matrix anti-commutes with the product

$$\gamma^5 = \gamma^0 \gamma^1 \gamma^2 \gamma^3. \tag{8.49}$$

We also need a matrix C which transforms γ^μ to its complex conjugate $\overline{\gamma}^\mu$, i.e.,

$$\gamma^m = C\,\overline{\gamma}^\mu C^{-1}.$$

Such a matrix exists since both γ^μ and $\overline{\gamma}^\mu$ will be realizations of the Clifford algebra for Minkowski space (which is real). In the Weyl representation (8.26),

$$C = \begin{pmatrix} 0 & i\,\sigma_y \\ -i\,\sigma_y & 0 \end{pmatrix}, \qquad i\,\sigma_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{8.50}$$

Now the operator of *charge conjugation* is the composition of C and complex conjugation. It leaves the Dirac matrices invariant, and transforms the wave function to its complex conjugate. Therefore it transforms the Dirac equation under an external field to the same equation, but with the sign of the charge e reversed, hence its name.

The discrete symmetries P, C, T played a key role in the development of relativistic quantum mechanics. It was originally thought that every system always admits all three as symmetries. However, in the early 1950's, it was experimentally discovered that the weak interaction does not conserve parity in general, so non-parity-symmetric equations such as the Weyl equation (8.47) for the neutrino are allowed. However, there is a general result, the *PCT Theorem*, that states that all theories must have the combined operator $P \circ C \circ T$ as a symmetry, even if they do not admit the individual operators.

Relativistic Spin

Note that the action of the Lorentz group $\mathrm{SO}(1,3)$, or, rather, $\mathrm{SL}(2,\mathbb{C})$, on a Dirac wave function ψ is given by the representation

$$\rho(A)\big[\psi(\mathbf{q})\big] = \rho_1(A)\,\psi(\rho_2(A)^{-1}\mathbf{q}),$$

where $\rho_1 = \rho^{1/2,0} \oplus \rho^{0,1/2}$ is the spin action, and $\rho_2 = \rho^{1/2,1/2}$ is the usual representation on Minkowski space. The corresponding conservation law for this action will consist of two pieces. For the rotation group $\mathrm{SU}(2)$, we have the usual angular momentum

$$J_\nu = x_\alpha\,\partial_\beta - x_\beta\,\partial_\alpha,$$

and the spin components

$$S_\nu = \tfrac{1}{4}\,[\gamma_\alpha, \gamma_\beta].$$

Note that $S_\nu = \tfrac{1}{2}\sigma_\nu$ when $\alpha \neq \beta$. Thus we re-derive the conservation of total angular momentum $\mathbf{L} = \mathbf{J} + \mathbf{S}$ for the Dirac equation. In particular, the electron spin of $\tfrac{1}{2}$ appears as it should.

Particle in an Electromagnetic Field

As with the non-relativistic theory, we obtain the corresponding equation for an electron or particle of charge e in a electromagnetic field with potential A_λ by replacing the ordinary momentum operator $P_\lambda = i\,\hbar\,\partial_\lambda$ by the perturbed operator

$$\widehat{P}_\lambda = i\,\hbar\,\partial_\lambda - e\,A_\lambda.$$

The Dirac equation in this case is

$$\left(\gamma^\lambda \widehat{P}_\lambda - m\right)\psi = 0. \tag{8.51}$$

Each component of ψ will then satisfy a scalar wave equation, which, however, is not obtained by replacing the momentum operators in the Klein-Gordon equation by their perturbed form! The spin will now appear, as follows. Operating on the above equation by $\gamma^\lambda \widehat{P}_\lambda + m$, we find

$$
\begin{aligned}
0 &= (\gamma^\lambda \widehat{P}_\lambda + m)(\gamma^\mu \widehat{P}_\mu - m)\psi = (\gamma^\lambda \gamma^\mu \widehat{P}_\lambda \widehat{P}_\mu + m^2)\psi \\
&= \left(\tfrac{1}{2}(\gamma^\lambda \gamma^\mu + \gamma^\lambda \gamma^\mu)\widehat{P}_\lambda \widehat{P}_\mu - \tfrac{1}{2}\gamma^\lambda \gamma^\mu [\widehat{P}_\lambda, \widehat{P}_\mu] + m^2\right)\psi \\
&= (g^{\lambda\mu}\widehat{P}_\lambda \widehat{P}_\mu + \mathrm{i}\,\hbar\, e\, S^{\lambda\mu} F_{\lambda\mu} + m^2)\psi,
\end{aligned}
$$

where $\mathbf{S}$ is the spin operator, and $F_{\lambda\mu} = \partial_\lambda A_\mu - \partial_\mu A_\lambda$ is the electromagnetic field; the electric components $\mathbf{E}$ correspond to the case $0 = \lambda < \mu$, the magnetic components $\mathbf{H}$ when $0 < \lambda < \mu$. Thus we recover the appropriate spin terms, which were added in somewhat ad hoc in the non-relativistic Schrödinger theory, directly from the Dirac equation!

Solutions of the Dirac Equation

We begin by considering the simplest solutions of the Dirac equation, which are plane waves (free particles), and are given by complex exponentials of the form

$$\psi(t,x) = \mathbf{u}\, e^{-\mathrm{i}\,\langle\, \mathbf{p}\,,\mathbf{q}\,\rangle}. \tag{8.52}$$

Here $\mathbf{p}$ is the relativistic momentum, $\mathbf{q}$ the position, and $\mathbf{u} \neq 0$ is a non-zero vector which must be determined by substituting into the Dirac equation. This requires that

$$(\gamma^\mu p_\mu - mc\,\mathrm{I})\,\mathbf{u} = 0, \tag{8.53}$$

hence the matrix $\gamma^\mu p_\mu - mc\,\mathrm{I}$ is not invertible. But

$$(\gamma^\mu p_\mu - mc\,\mathrm{I})(\gamma^\mu p_\mu + mc\,\mathrm{I}) = |\,\mathbf{p}\,|^2 - m^2 c^2 = p^2 - m^2 c^2,$$

hence the matrix is not invertible if and only if

$$p^2 = m^2 c^2, \qquad i.e., \qquad E^2 = c^2 p^2 + m^2 c^4,$$

which shows that the relativistic formula for the energy has been preserved. The Dirac equation has both positive and negative energy solutions, corresponding to

$$E = \pm\sqrt{c^2 p^2 + m^2 c^4}\,.$$

Each of these is doubly degenerate because of the charge conjugation (spin in direction of the momentum) symmetry. Therefore the eigenvalues of the Hamiltonian for this value of the energy are classified by the sign of the energy $\pm E$ and the sign of the spin $\pm\tfrac{1}{2}$.

If we use the Dirac representation (8.25), then the explicit solution to (8.53) when the momentum is parallel to the z axis is a linear combination of the solutions

$$\psi_{++} = \begin{pmatrix} 1 \\ 0 \\ \alpha \\ 0 \end{pmatrix}, \qquad \psi_{+-} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -\alpha \end{pmatrix}, \qquad \psi_{-+} = \begin{pmatrix} -\alpha \\ 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \psi_{--} = \begin{pmatrix} 0 \\ \alpha \\ 0 \\ 1 \end{pmatrix},$$

where

$$\alpha = \frac{cp}{E + mc^2} = \sqrt{\frac{E - mc^2}{E + mc^2}}.$$

The first two have positive energy, indicated by the first $+$; the second sign indicates the sign of the spin. The normalization factor for all these solutions is $\sqrt{2E/(E + mc^2)}$. In the positive energy solutions, the first two components are called the "large components", and the second two "small components", since, in the non-relativistic limit, α is small.

Spherically Symmetric Potential

The motion of an electron in an atom is a special case of the motion of a particle in a spherically symmetrical potential. The Dirac equation in this case is

$$\left[\gamma^\mu P_\mu - mc + V(r) \right] \psi = 0. \tag{8.54}$$

As in the non-relativistic Schrödinger equation, one looks for solutions which are joint eigenfunctions of the angular momentum operators J^2, J_z, and the parity reversal P, i.e.,

$$J^2 \psi = j(j+1)\,\psi, \qquad J_z \psi = l\,\psi, \qquad P\,\psi = (-1)^{j + \varpi/2}\,\psi,$$

where j is the angular momentum quantum number, l the magnetic quantum number, and $\varpi = \pm 1$ indicates the parity of the state. Separating variables in the Dirac equation leads to a solution of the form

$$\psi(t, r, \varphi, \theta, t) = \frac{e^{\,\mathrm{i}\,Et/c}}{r} \begin{pmatrix} \alpha(r)\,\mathbf{Y}(\varphi, \theta) \\ \beta(r)\,\widetilde{\mathbf{Y}}(\varphi, \theta) \end{pmatrix}, \tag{8.55}$$

where E is the energy, $\mathbf{Y}(\varphi, \theta), \widetilde{\mathbf{Y}}(\varphi, \theta)$ are suitable spherical harmonics combined with spin components for the two representations $\rho^{1/2,0}$ and $\rho^{0,1/2}$ respectively. The functions $\alpha(r), \beta(r)$ satisfy the system of first order ordinary differential equations

$$\begin{aligned} -\hbar\,\alpha' + \frac{\varpi(j + \tfrac{1}{2})}{r}\,\alpha &= \left[\frac{E}{c} - mc - V(r) \right] \beta, \\ \hbar\,\beta' + \frac{\varpi(j + \tfrac{1}{2})}{r}\,\beta &= \left[\frac{E}{c} + mc - V(r) \right] \alpha, \end{aligned} \tag{8.56}$$

For $V(r) \equiv 0$ (i.e., a free particle), the solution to this system can be expressed in terms of Bessel functions. For the Coulomb potential $V(r) = -Ze^2/r$, a formal series solution takes the form

$$\alpha = r^s e^{-\kappa r}(a_0 + a_1 r + \cdots), \qquad \beta = r^s e^{-\kappa r}(b_0 + b_1 r + \cdots), \tag{8.57}$$

where, from the indicial equation,

$$\kappa = \sqrt{\frac{m^2 c^4 - E^2}{c^2}}, \qquad s = \sqrt{\left(j + \tfrac{1}{2}\right)^2 - Z^2 e^4}.$$

As with the non-relativistic Schrödinger equation, it can be proved that the series (8.57) will correspond to a bound state if and only if it terminates. Calculating the recursion relation for the coefficients, we find that this happens precisely at the energy levels

$$\hbar^2 E_{n,j} = \frac{m c^2}{\sqrt{1 + \dfrac{Z^2 e^4}{\left(n + s - j - \tfrac{1}{2}\right)^2}}}. \tag{8.58}$$

If we expand this in inverse powers of n, we find

$$E_{n,j} = \frac{m c^2}{\hbar^2} \left[1 - \frac{Z^2 e^4}{2 n^2} - \frac{Z^4 e^8}{2 n^4} \left(\frac{n}{j + \tfrac{1}{2}} - \frac{3}{4} \right) + \cdots \right]. \tag{8.59}$$

The first term is the relativistic rest mass of the electron, which must be ignored in any non-relativistic calculation of energy levels. The second term coincides with the non-relativistic formula for the energy levels of the hydrogen atom; in particular it retains the accidental degeneracy in that it does not depend on the orbital angular momentum quantum number j. The following term is the relativistic correction to the energy levels. Note that the energy level is slightly decreased, with the larger decrease happening for smaller angular momenta j. This agrees broadly with the experimental results on the fine structure of the spectrum. The agreement, however, is not perfect, and additional corrections, to account for the so-called Lamb shift, are based on "radiative corrections" taking into account the interactions between the electron, the nucleus, and the quantized electromagnetic field.

Non-relativistic Limit

In the non-relativistic limit, the energy E differs little from the relativistic rest energy, i.e., the kinetic energy K satisfies

$$K = E - m c^2 \ll m c^2.$$

This implies that the velocity of the particle is small:

$$\frac{K}{m} = \tfrac{1}{2} v^2 \ll 1.$$

In this case the small components of the wave function ψ have size $\mathrm{O}(v/c)$.

The easiest way to compute the non-relativistic limit is with the Dirac form (8.25) of the Dirac matrices, where

$$\gamma_0 \overset{\longrightarrow}{\Longrightarrow} \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \qquad \gamma_k \overset{\longrightarrow}{\Longrightarrow} \begin{pmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{pmatrix}.$$

The relativistic energy of the electron will includes its relativistic rest energy mc^2, which must be excluded to arrive at the non-relativistic approximation. Thus we set

$$\psi = \begin{pmatrix} \varphi \\ \chi \end{pmatrix} \exp\left(-\frac{\mathrm{i}\,mc^2t}{\hbar} \right)$$

for the wave function, where φ is the large and χ the small component. Substituting into the Dirac equation, we deduce a system of two two-component equations

$$\mathrm{i}\,\hbar\,\partial_t\varphi + c\,\sigma^k P_k\,\chi = 0, \qquad (\mathrm{i}\,\hbar\,\partial_t + 2\,mc^2)\chi + c\,\sigma^k P_k\,\varphi = 0, \qquad (8.60)$$

which are so far identical with the Dirac equation. Now in the second system, only the term $2mc^2\chi$ is comparable in magnitude to φ, so we can neglect the first term, solve for χ and substitute into the first system. This gives

$$\mathrm{i}\,\hbar\,\partial_t\varphi = \frac{1}{2\,m}(\sigma^k P_k)^2\varphi.$$

However, using the identity

$$(\sigma^k a_k)\,(\sigma^k b_k) = a \cdot b + \mathrm{i}\,\boldsymbol{\sigma} \cdot a \wedge b,$$

we deduce that the system reduces to

$$\mathrm{i}\,\hbar\,\partial_t\varphi = -\frac{1}{2\,m}P_k P^k\,\varphi,$$

i.e., it decouples into a pair of scalar Schrödinger equations.

If we perform the same analysis for an electron in a magnetic field, with four-potential $\mathbf{A} = (\mathbf{E}, \mathbf{B})$, then an additional term due to the cross product in the above identity enters the equation:

$$\mathrm{i}\,\hbar\,\partial_t\varphi = \left[\frac{1}{2\,m}\left(P_k - \frac{e}{c}B_k \right)^2 + eE - \frac{e\hbar}{2\,mc}\sigma^k(\mathrm{curl}\,\mathbf{B})_k \right]\varphi. \qquad (8.61)$$

The additional term coming from the curl of the magnetic potential is exactly the terms introduced ad hoc earlier to account for the electron spin. Thus the Dirac theory, in the non-relativistic limit, reduces to the Pauli theory!

A significant problem is to figure out what to do with the negative energy states predicted by the Dirac equation. Dirac postulated that all these states are full, and so, by the Pauli exclusion principle, no positive energy states can cross over. This is referred to as the "Dirac sea". However, this does allow the possibility of a negative energy wave function becoming one with positive energy, leaving behind a hole in the Dirac sea. These holes correspond to positrons, but were initially discounted by Dirac due to lack of experimental evidence. The theory was dramatically vindicated after their discovery the following year.

References

[**1**] Ablowitz, M.J., and Segur, H., *Solitons and the Inverse Scattering Transform*, Stud. Appl. Math., vol. 4, SIAM, Philadelphia, 1981.

[**2**] Abraham, R., and Marsden, J.E., *Foundations of Mechanics*, 2nd ed., The Benjamin–Cummings Publ. Co., Reading, Mass., 1978.

[**3**] Adams, R.A., *Sobolev Spaces*, Academic Press, New York, 1975.

[**4**] Akin, K., Buchsbaum, D. A., and Weyman, J., Schur functors and Schur complexes, *Adv. Math.* **44** (1982), 207–278.

[**5**] Arnol'd, V.I., *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics, vol. 60, Springer–Verlag, New York, 1978.

[**6**] Atiyah, M.F., Bott, R., and Shapiro, A., Clifford modules, *Topology* **3**, Suppl. 1 (1964), 3–38.

[**7**] Berestetskii, V.B., Lifschitz, E.M., and Pitaevskii, L.P., *Quantum Electrodynamics*, 2nd ed., Pergamon Press, New York, 1982.

[**8**] Bethe, H.A., and Jackiw, R., *Intermediate Quantum Mechanics*, 3rd ed., Benjamin/Cummings Publ. Co., Menlo Park, CA, 1986.

[**9**] Biedenharn, L.C., and Louck, J.D., *Angular Momentum in Quantum Physics*, Addison–Wesley, Reading, Mass., 1981.

[**10**] Born, M., and Wolf, E., *Principles of Optics*, 4th ed., Pergamon Press, New York, 1970.

[**11**] Bott, R., Homogeneous vector bundles, *Ann. Math.* **66** (1957), 203–248.

[**12**] Butler, P.H., and Wybourne, B.G., Tables of outer S-function plethysms, *Atomic Data* **3** (1971), 133–151.

[**13**] Carter, R.W., and Payne, J.T., On homomorphisms between Weyl modules and Specht modules, *Math. Proc. Camb. Phil. Soc.* **87** (1980), 419–425.

[**14**] Chadan, K., and Sabatier, P.C., *Inverse Problems in Quantum Scattering Theory*, Springer–Verlag, New York, 1977.

[**15**] Chen, J.Q., *Group Representation Theory for Physicists*, World Scientific, Singapore, 1989.

[**16**] Chevalley, C., *The Algebraic Theory of Spinors*, Columbia University Press, New York, 1974.

[**17**] Courant, R., and Hilbert, D., *Methods of Mathematical Physics*, vol. II, Interscience Publ., New York, 1953.

[**18**] Crumeyrolle, A., *Orthogonal and Symplectic Clifford Algebras*, Kluwer Acad. Publ., Boston, 1990.

[**19**] Curtis, C.W., and Reiner, I., *Representation Theory of Finite Groups and Associative Algebras*, Interscience, New York, 1962.

[**20**] Dirac, P.A.M., *The Principles of Quantum Mechanics*, 3rd ed., Clarendon Press, Oxford, 1947.

[**21**] Dodd, R.K., Eilbeck, J.C., Gibbon, J.D., and Morris, H.C., *Solitons and Nonlinear Wave Equations*, Academic Press, New York, 1982.

[**22**] Feynman, R.P., *The Character of Physical Law*, M.I.T. Press, Cambridge, MA, 1965.

[**23**] Frame, J.S., Robinson, G. de B., and Thrall, R.M., The hook graphs of S_n, *Can. J. Math.* **6** (1954), 316–324.

[**24**] Gel'fand, I.M., and Dikii, L.A., Asymptotic behavior of the resolvent of Sturm-Liouville equations and the algebra of the Korteweg–deVries equation, *Russian Math. Surveys* **30**:5 (1975), 77–113; also *Collected Papers*, vol. 1, Springer-Verlag, New York, 1987, pp. 647–683.

[**25**] Gel'fand, I.M., and Fomin, S.V., *Calculus of Variations*, Prentice–Hall, Inc., Englewood Cliffs, N.J., 1963.

[**26**] Glimm, J., and Jaffe, A., *Quantum Physics*, 2nd ed., Springer–Verlag, New York, 1987.

[**27**] Goldstein, H., *Classical Mechanics*, 2nd ed., Addison–Wesley, Reading, Mass., 1980.

[**28**] Guillemin, V., and Sternberg, S., *Geometric Asymptotics*, American Mathematical Society, Providence, R.I., 1977.

[**29**] Heine, V., *Group Theory in Quantum Mechanics*, Pergamon Press, New York, 1960.

[**30**] Landau, L.D., and Lifshitz, E.M., *Quantum Mechanics* (*Non-relativistic Theory*), Course of Theoretical Physics, vol. 3, Pergamon Press, New York, 1977.

[**31**] Lax, P.D., Integrals of nonlinear equations of evolution and solitary waves, *Commun. Pure Appl. Math.* **21** (1968), 467–490.

[**32**] Luneburg, R.K., *Mathematical Theory of Optics*, University of California Press, Berkeley, 1966.

[**33**] Macdonald, I.G., *Symmetric Functions and Hall Polynomials*, Clarendon Press, Oxford, 1979.

[**34**] Maliakis, M., and Olver, P.J., Explicit generalized Pieri maps, *J. Algebra* **148** (1992), 68–85.

[**35**] Maslov, V.P., and Fedoryuk, M.V., *Semi-classical Approximation in Quantum Mechanics*, D. Reidel Publishing Co., Boston, 1981.

[**36**] Messiah, A., *Quantum Mechanics*, John Wiley & Sons, New York, 1976.

[**37**] Miller, W., Jr., *Symmetry Groups and Their Applications*, Academic Press, New York, 1972.

[**38**] Naimark, M. A., *Linear Representations of the Lorentz Group*, Macmillan, New York, 1964.

[**39**] Naimark, M.A., and Stern, A.I., *Theory of Group Representations*, Springer–Verlag, New York, 1982.

[**40**] Olver, F.W.J., *Asymptotics and Special Functions*, Academic Press, New York, 1974.

[**41**] Olver, F.W.J., Lozier, D.W., Boisvert, R.F., and Clark, C.W., eds., *NIST Handbook of Mathematical Functions*, Cambridge University Press, Cambridge, 2010.

[**42**] Olver, P.J., *Applications of Lie Groups to Differential Equations*, Second Edition, Graduate Texts in Mathematics, vol. 107, Springer–Verlag, New York, 1993.

[**43**] Olver, P.J., *Introduction to Partial Differential Equations*, Undergraduate Texts in Mathematics, Springer, New York, 2014.

[**44**] Olver, P.J., and Shakiban, C., *Applied Linear Algebra*, Second Edition, Undergraduate Texts in Mathematics, Springer, New York, 2018.

[**45**] Reed, M., and Simon, B., *Methods of Modern Mathematical Physics*, vol. 1, Academic Press, New York, 1972.

[**46**] Richtmyer, R.D., *Principles of Advanced Mathematical Physics*, vol. 1, Springer-Verlag, New York, 1978.

[**47**] Schrödinger, E., *Collected Papers on Wave Mechanics*, Chelsea Publ. Co., New York, 1982.

[**48**] Scott, A.C., Chu, F.Y.F., and McLaughlin, D.W., The soliton: a new concept in applied science, *Proc. IEEE* **61** (1973), 1443–1482.

[**49**] Talman, J.D., *Special Functions. A Group Theoretic Approach*, W.A. Benjamin, New York, 1968.

[**50**] Thirring, W., *Quantum Mechanics of Atoms and Molecules*, A Course in Mathematical Physics, vol. 3, Springer–Verlag, New York, 1981.

[**60**] Wong, R., *Asymptotic Approximation of Integrals*, Academic Press, New York, 1989.

[**52**] Titchmarsh, E. C., *Fourier Series and Integrals*, Oxford University Press, London, 1948.

[**53**] van der Waerden, B.L., *Group Theory and Quantum Mechanics*, Springer–Verlag, New York, 1980.

[**54**] Weyl, H., *Classical Groups*, Princeton University Press, Princeton, N.J., 1946.

[**55**] Weyl, H., *The Theory of Groups and Quantum Mechanics*, Dover Publ., New York, 1950.

[**56**] Weyman, J., *Cohomology of Vector Bundles and Syzygies*, Cambridge University Press, Cambridge, 2003.

[**57**] Whitham, G.B., *Linear and Nonlinear Waves*, John Wiley & Sons, New York, 1974.

[**58**] Whittaker, E.T., *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge, 1937.

[**59**] Woodhouse, N.M.J., *Geometric Quantization*, Second edition; Oxford University Press, New York, 1992.

[**60**] Wong, R., *Asymptotic Approximation of Integrals*, Academic Press, New York, 1989.

[**61**] Wybourne, B.G., *Symmetry Principles and Atomic Spectroscopy*, Wiley–Interscience, New York, 1970.